Spring 4-17-2020

E.coli in Tanyard Creek: An Analysis on the Temporal and Spatial Patterns and the Influence of Rainfall

Anam Syed

ABSTRACT

***E. COLI* IN TANYARD CREEK: AN ANALYSIS ON THE TEMPORAL AND SPATIAL PATTERNS AND THE INFLUENCE OF RAINFALL.**

By

ANAM SYED

MAY 2020

INTRODUCTION: Tanyard Creek is an urban creek in the City of Atlanta. It receives the treated discharge from Tanyard Creek Combined Sewage Overflow (CSO), before flowing into the Chattahoochee River. One of the significant issues surrounding the creek is the lack of regulation, monitoring, and maintenance -- leading to broken pipes, fecal pollution and input of street debris According to the recreational water quality standards, Tanyard Creek is impaired because of the high concentration of *E. coli* and other microbials present in the creek.

AIM: This study will analyze, compare, and observe *E. coli* concentration present in Tanyard Creek over time (October 2018-February 2020) and spatially (from site to site). Additionally, this study will determine if there is a relationship between rainfall amounts in Atlanta and elevated *E. coli* concentration in Tanyard Creek over the periods 24 and 48 hours before sampling, as well as cumulatively.

METHODS: Water samples were collected from Tanyard Creek on a weekly basis for 54 weeks. For this research, the sampling time duration is from October 16, 2018, through February 12, 2020. Samples were collected from 10 pre-determined sites every week, and all the sites were located along the stretch of a 1-mile trail. The samples were processed using membrane filtration and BioRad RAPID'*E. coli* 2™ medium was used for the assay of *E. coli* count.

RESULTS AND DISCUSSION: Temporally, there is variability in the *E. coli* concentration on a weekly basis, and there is no particular trend by seasons. The results support the hypothesis, and *E. coli* concentrations in Tanyard Creek consistently violate EPA recreational water quality standards and the creek is impaired for recreational use. For all three of the rainfall amounts, there was no relationship between rainfall and elevated *E. coli* concentration. The only consistent factor is that the average *E. coli* concentrations exceed the acceptable EPA standards for recreational waters.

E. COLI IN TANYARD CREEK: AN ANALYSIS ON THE TEMPORAL AND SPATIAL PATTERNS AND THE INFLUENCE OF RAINFALL.

by

ANAM SYED

B.S., GEORGIA STATE UNIVERSITY

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

E. COLI IN TANYARD CREEK: AN ANALYSIS ON THE TEMPORAL AND SPATIAL PATTERNS AND THE INFLUENCE OF RAINFALL.

by

ANAM SYED

Approved:

____
Lisa Casanova, Ph.D.
Committee Chair

____
Christine Stauber, Ph.D.
Committee Member

Defense Date: April 17, 2020

Author’s Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

____

Anam Syed

TABLE OF CONTENTS

Acknowledgements

My sincerest gratitude to my thesis committee chair, Dr. Lisa Casanova for the opportunity to work on this project. Thank you for the guidance, wisdom, knowledge and mentorship throughout my journey in this program. My appreciation also extends to Dr. Christine Stauber for joining my thesis committee and for her guidance throughout the thesis writing process. I would also like thank Michael Kshatri and Aaron Murihead, the two most wonderful lab technicians and friends. Your guidance, encouragement and help has been essential throughout this project. Thank you, Katie Pearce and Hannah-Leigh Crawford, my fellow lab partners, who have kept me motivated throughout the writing process, and have provided constant feedback and ideas for this project. Lastly, I would like to thank my friends and family for their constant love and support throughout my journey.

List of Tables

List of Figures

Chapter I: Introduction

1.1 Recreational Water Quality

Recreational waters are a source of gastrointestinal disease outbreaks that require public health attention (U.S., EPA 2012). Surface recreational waters such as coastal beaches, rivers, and lakes can potentially be contaminated with fecal indicator bacteria (FIB) such as *Escherichia coli (E. coli)* and *Enterococci.* The presence of *E. coli* suggests that human and/ or animal feces are present in the water, because *E. coli* is an enteric bacterium that is naturally found in humans and other warm-blooded animals (Schilling et al., 2009). It also poses the threat of potential health risks such as gastrointestinal illnesses (GI illness), especially in immunocompromised and vulnerable people in contact with the water. Input of FIB can come from direct deposition, point sources, diffuse sources, and resuspension of FIB contained in sediments. Animal fecal sources can contaminate recreational bodies of water through direct fecal loading water, and indirect contamination can occur via runoff from the land. Finally, the changing environmental conditions (for example, heavy rainfall events, weather) can also play a role in elevated FIB in recreational waters (U.S., EPA, 2012).

If waters violate the EPA recreational water quality standards consistently, then states are required to categorize them as "impaired." The current EPA standards for recreational waters is a geometric mean of 126 Colony Forming Units (CFU) of *E. coli* per 100 mL of water (U.S., EPA 2012). There is also a standard for the Total Maximum Daily Load (TMDL) of pollutants that a waterbody can support while serving its designated use (U.S., EPA 2018). If recreational waters violate one or both of these standards, then they will be categorized as impaired. A list of impaired recreational waters is compiled in Georgia every two years, but the requirement to comply with water quality standards is not stringent enough for some bodies of water.

1.2 Combined Sewage Overflow in Atlanta

In 1998, there was a lawsuit filed by the Georgia Department of Natural Resources, Environmental Protection Division (GA EPD), and three citizen plaintiffs against the City of Atlanta to eliminate water quality violations. The City violated the Clean Water Act and Georgia Water Quality Control Act resulting from the discharge of effluent from Combined Sewage Overflow (CSO) facilities into urban streams that empty into the Chattahoochee River. The City of Atlanta paid a $3.2 million fine for the violations. The second consent decree was issued in 1999 to address sanitary sewer problems. By July of 2001, the federal EPA and GA EPD approved the City's plan to eliminate water quality violations from CSO. The plan consisted of the following goals:

1. Construct a deep-rock tunnel storage and treatment system to capture and store combined stormwater and sewage flow to two new CSO treatment facilities for treatment before discharge to the Chattahoochee or South Rivers.
2. Screen the overflows while disinfecting and dechlorinating them prior to discharge, so they meet the water quality standards.

CSOs carry wastewater from residential and commercial buildings. When they overflow because of heavy rainfall or excess stormwater, all the discharge from the wastewater flows directly into the Chattahoochee or South Rivers. The discharge contains high levels of bacteria and pollutants that are potential environmental and health risks. The City proposed a sewer system to separate the sanitary water from the stormwater. The sanitary waters consist of residences and businesses, and all the raw waste and other defects produced are discharged into the stormwater system. The stormwater is then fully treated in the reclamation centers before being discharged into the Chattahoochee or South Rivers (**Figure 1.21**). However, if rainfall overwhelms the capacity of the reclamation centers, the stormwater is diverted to a CSO with minimal treatment. If rainfall further causes an overflow at the CSO, then the wastewater is discharged into its designated creek with minimal treatment. This treatment entails wastewater disinfection with hypochlorite (chlorine) that violates water quality standards.

The CSO projects were finished in October of 2008, resulting in seven CSOs around the City of Atlanta **(Figure 1.22)** and cost approximately $711 million. In January of 2014, a Performance Audit was completed to assess the Department of Watershed Management's (DWM) efforts to comply with the 1998 CSO consent decree. The results were as follows: decrease in water quality violations by 65% from 2008-2013, the severity of violations decreased after implementing improvements, and exceeding total fecal coliform bacteria into the Chattahoochee and South Rivers decreased from 74% to 43%. However, issues of street-level debris clogging sewer systems, damaging filtering equipment, and introducing new pollutants into the CSO system persisted. By December of 2011, equipment such as water mains, catch basins, pump stations, tanks, and water meters were beginning to wear out and needed replacement because they were causing water quality violations. In summary, the CSO facilities had no preventative maintenance. In order to continue improving the City's CSO facilities, an ongoing budget to cover future maintenance needs and cost-effective methods need to be implemented.

1.3 Tanyard Creek

Tanyard Creek is an urban creek in the City of Atlanta. It receives the discharge from Tanyard Creek CSO, which is one of the West Area CSOs. From 1998-2013, these West Area CSOs had the highest number of water quality violations. They required higher levels of treatment compared to other CSOs, including removal of solids, metals, filtration, and disinfection by UV to destroy bacteria and viruses. Higher levels of treatment now take place in the West Area Tunnel that stores approximately 177 million gallons of overflow from the Clear Creek, Tanyard Creek, and North Avenue CSO facilities. Once the wastewater is treated, it is then released into the designated creek, followed by discharge into the Chattahoochee River **(Figure 1.31).**

Tanyard Creek receives treated overflow water from the Tanyard Creek CSO in accordance with the consent decree. There are yellow signs around Tanyard Creek, warning the public not to play, swim, fish in the creek (**Figure 1.32)** because it is subject to sewage overflows and runoff contaminants. During the summer months, there is often an unpleasant odor emanating from the creek and pollutants overflowing around the railroad bridge. Tanyard Creek is considered to be impaired, because it consistently fails to meet the TMDL cutoff for its designated use (GA EPD 305(b)/303(d) Draft, 2020). One of the significant issues surrounding the creek is the lack of regulation, monitoring, and maintenance -- leading to broken pipes, fecal pollution and input of street debris. In order to assess the fecal contamination, and magnitude of impairment in Tanyard Creek, this research will monitor *E. coli* patterns spatially (from site to site) and over time.

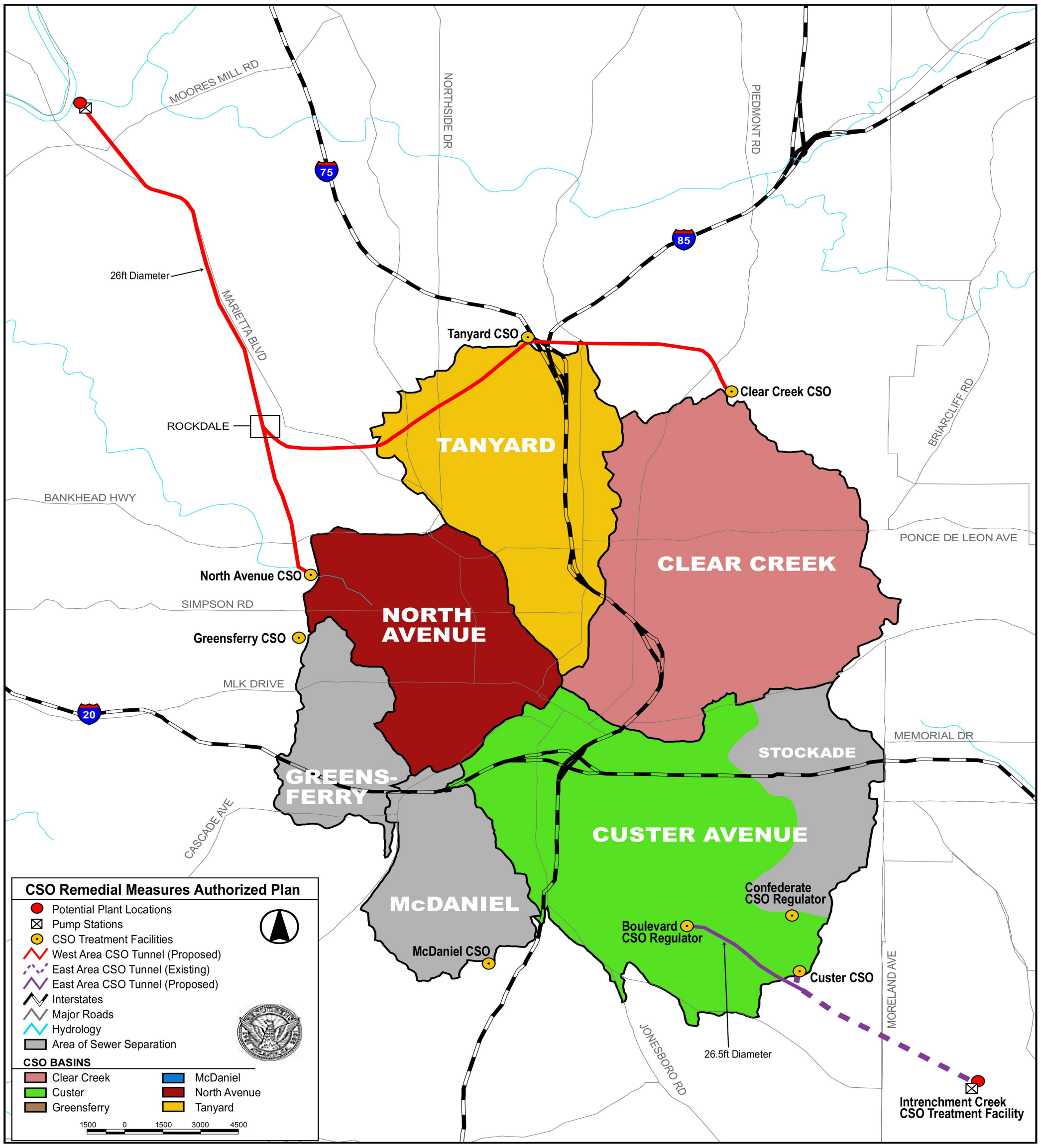
1.4 Research Question, Aims, and Hypothesis

Research Question: How does *E. coli* concentration vary spatially (from site to site) and over time (October 2018-February 2020) in Tanyard Creek? Are rainfall quantities 24 hours, 48 hours, and 48 hours cumulative prior to sampling associated with elevated *E. coli* concentration in Tanyard Creek?

Aim 1: Measure *E. coli* concentration present in Tanyard Creek over time (October 2018-February 2020) and spatially (from site to site).

Aim 2: Examine possible relationships between rainfall amounts in Atlanta and elevated *E. coli* concentration in Tanyard Creek 24 and 48 hours before sampling as well as cumulatively (48 hours cumulative).

Figure 1.21. *Separating sanitary water and stormwater*. Taken from: Performance Audit: Combine Sewage Overflow Consent Decree Impact (January 2014). City Auditors Office, City of Atlanta. **Source:** Cambridge Department of Public Works, Cambridge Massachusetts.



CSO Master Plan
Option 1 - Separate McDaniel, Greensferry, and Stockade

082002011ATL/CSO 101.ai

Figure 1.22 *Map of the CSO facilities in Atlanta.*

Exhibit 2 Water Flow and Treatment at the West Area CSO Facilities

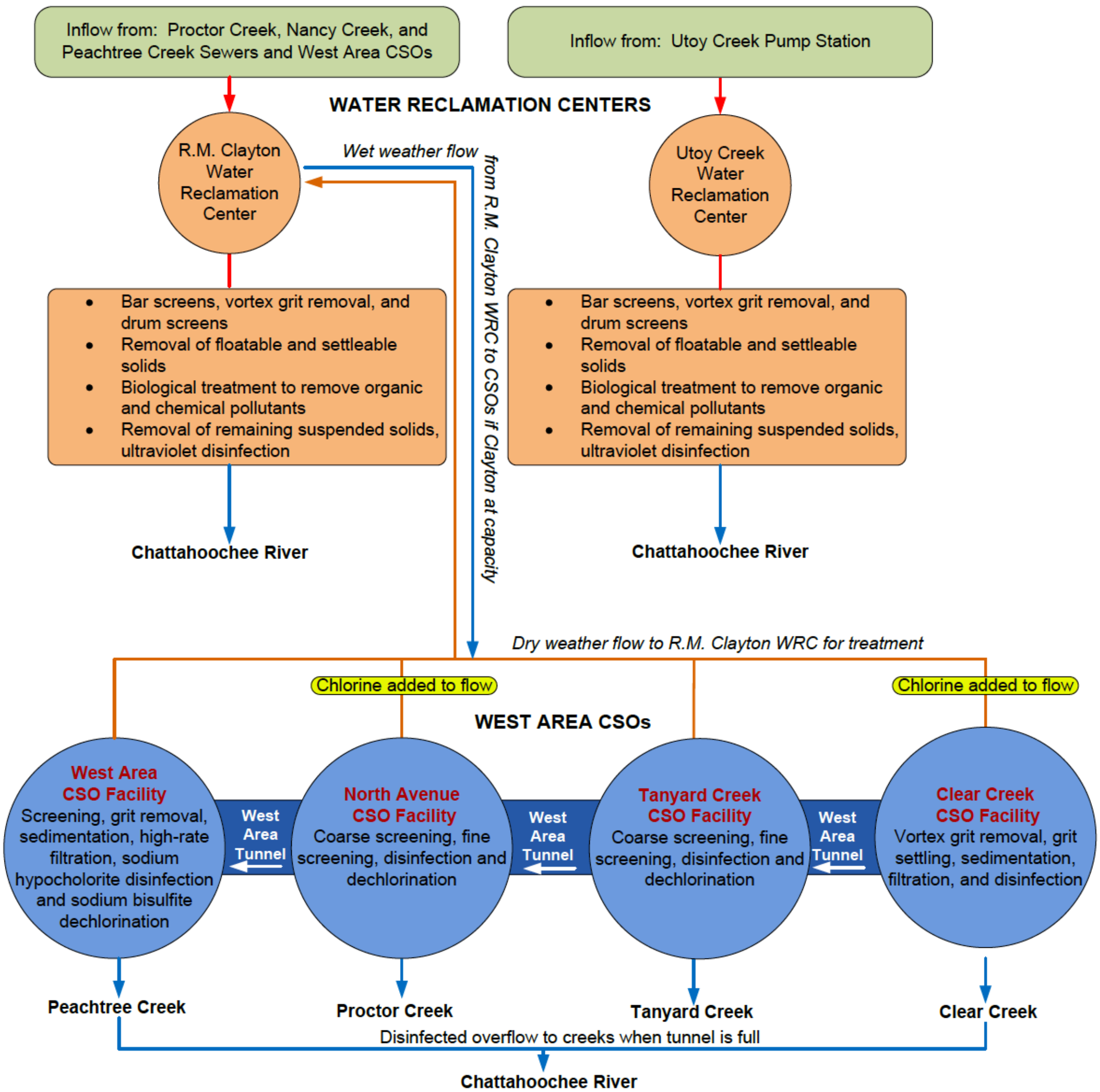


Figure 1.31. *Water flow and treatment of the West Area CSO Facilities.* Taken from: Performance Audit: Combine Sewage Overflow Consent Decree Impact (January 2014). City Auditors Office, City of Atlanta. **Source:** Prepared by city auditor's staff using information from the Department of Watershed Management

Figure 1.32 *Warning Sign at Tanyard Creek*

Chapter II: Literature Review

2.1 Combined Sewage Overflow (CSO)

According to the Miami Conservancy District Report of 2018, stormwater that originates in urban areas is a major source of pollution. Development and urbanization have increased the detrimental impact of stormwater by increasing the volume of runoff and non-natural input into the larger bodies of water (ex. Great Lakes and the Chattahoochee River) (Fisher et al., 2015). Urban runoff is responsible for 32% of water quality impairment in estuaries, which flow into the Great Lakes in Michigan (U.S. EPA, 2012). The wastewater is usually subjected to fecal contamination from point and non-point sources; point sources include combined sewage overflows (CSO), agricultural runoff, urban stormwater, and streams. There are also pollutants such as oil, grease, pesticides, chemicals, and more incoming from streets, as well as parking lots in residential and commercial areas (Kleinheinz et al., 2009).

Urban stormwater delivers runoff to surface waters during heavy rain events because of failing infrastructure, and CSOs are at the core of the issue. CSOs collect and convey stormwater and wastewater through a single pipe network, and during heavy rainfall this water is discharged into local surface waters. Therefore, CSOs play an integral role in urban water pollution, and their discharges cause water impairment and establish cause for a public health concern (Lund et al., 2014). CSOs serve as point sources that foster the growth of microbial organisms that would otherwise have low abundance in the natural environment. These microorganisms grow on surfaces such as pipes, roads, as well as soil and are discharged directly into local creeks and rivers (Fisher et al., 2015). The microbial contents and exposure pathogens that originate from CSOs are associated with increased health risks such as waterborne diseases and gastrointestinal illnesses (GI Illnesses).

2.2 *E. coli* as a Fecal Indicator Bacteria

Urban ecology addresses the interactions of organisms and the environment within built landscapes and bacteria plays an essential role in the ecosystem by providing nutrient cycling and pollution degradation (Fisher et al., 2015). *E. coli* is an enteric bacterium and is naturally found in humans and other warm-blooded animals. High concentrations of *E. coli* in water indicate the presence of human and/or fecal animal waste, and the possible existence of other pathogens (Schilling et al., 2009). According to the Recreational Water Summary of 2012, previous epidemiological studies concluded that *E. coli* consistently performed well as indicators of illness in sewage-contaminated waters during studies for freshwater. Moreover, epidemiological studies have also concluded *E. coli* is associated with GI illnesses from exposure to recreational waters. Not only does *E. coli* indicate the presence of waste in a waterbody, but elevated *E. coli* in recreational water is correlated with GI Illnesses (r=0.804) (Rossi et al., 2020).

E. coli are Fecal Indicator Bacteria (FIB) that are monitored in recreational waters in order to decrease the risk of GI illnesses (Benjamin-Chung et al., 2017). The Clean Water Act and the Beach Environmental Assessment and Coastal Health (BEACH) Act of 2000 made it mandatory to monitor the FIB present in recreational waters and warn the public if it exceeds acceptable standards. According to the 1986 criteria by the USEPA: Exceedances of the standards may result in a state listing the waterbody as impaired in accordance with the Clean Water Act (CWA). Listing a waterbody as impaired initiates the Total Maximum Daily Load (TMDL) process which ultimately leads to implementing remediated actions, so that water quality standards can be met and designated uses, such as recreational water contact, can be attained (Boehm et al., 2009). The current EPA standards for *E. coli* in recreational waters is 126 colony forming units (CFU) per 100mL of water **(Table 2.21)** (Rossi et al, 2020).

2.3 Rainfall and *E. coli*

There are currently recreational waters that are deemed impaired because of high *E. coli* concentration, an example being the Raccoon River in Iowa. A study was done over eight years to determine what conditions cause *E. coli* to exceed regulatory standards, and rainfall was one of the possible factors (Schilling et al., 2009). The study found that *E. coli* concentrations were highest during months of high rainfall (May-July), but the high concentration could also be attributed to storm runoff and impact from point sources. Another conclusion from this study was that *E. coli* and rainfall correlated over a short period (a few days), and it was harder to find a long-term persistence in concentration. Rossi et al., conducted a similar study that concluded cumulative rainfall 72 hours prior to sampling elevated *E. coli* concentration significantly (at alpha=0.05). The elevation caused by rainfall increases the flow discharge and turbulence of water, causing resuspension of sediments and pathogens in the river. The same conclusion was made by the Miami Conservancy District Report of 2018 when comparing wet versus dry weather. The effect of rainfall is also substantial enough that county health departments in Southern California issue warnings for the public to avoid recreational water contact for three days following storms, which result in >25mm of rainfall. Every storm that resulted in >25mm of rain caused an elevated bacterial concentration that failed the EPA standards for acceptable *E. coli* in recreational waters. The same results were concluded for rainfall between 6-25mm (Ackerman & Weisberg, 2003).

The relationship between rainfall and elevated *E. coli* concentrations are contrasted with studies that concluded that there is no relationship between the two. A study conducted by Kleinheinz et al., concluded that rainfall more than 2.5mm resulted in higher microbial contamination, and rainfall less than that amount had no significant impact. In the same study by Ackerman & Weisberg in Southern California, there was no uniform correlation between rainfall 2.5-6mm and elevated *E. coli* concentration. Finally, there is the hypothesis of changing environmental conditions influencing sediment resuspension and causing elevated FIB in water, but there is no definite conclusion that rainfall is the factor (U.S. EPA, 2009).

Table 1. Recommended 2012 RWQC.

Criteria Elements	**Estimated Illness Rate (NGI): 36 per 1,000 primary contact recreators**			**Estimated Illness Rate (NGI): 32 per 1,000 primary contact recreators**	
	Magnitude			**Magnitude**	
Indicator	GM (cfu/100 mL)[a]	STV (cfu/100 mL)[a]	OR	GM (cfu/100 mL)[a]	STV (cfu/100 mL)[a]
Enterococci – marine and fresh	35	130		30	110
OR					
E. coli – fresh	126	410		100	320
Duration and Frequency: The waterbody GM should not be greater than the selected GM magnitude in any 30-day interval. There should not be greater than a ten percent excursion frequency of the selected STV magnitude in the same 30-day interval.					

[a] EPA recommends using EPA Method 1600 (U.S. EPA, 2002a) to measure culturable enterococci, or another equivalent method that measures culturable enterococci and using EPA Method 1603 (U.S. EPA, 2002b) to measure culturable *E. coli*, or any other equivalent method that measures culturable *E. coli*.

Table 2.21 *The 2012 Recreational Water Quality Criteria for acceptable E. coli CFU (126CFU/100mL =2.1 log_{10}CFU/100mL)*

Chapter III: Methods

3.1 Primary Data Collection

Water samples were collected from Tanyard Creek on a weekly basis. The samples were collected over a span of 16 months, from October 2018 to February 2020. Weeks in which samples were not collected were excluded from the analysis. Duplicates were collected from each site in a sterile sample bag. Samples were collected from the same 10 pre-determined sites every week, and all ten samples were collected on the same day. The sites were all located along the stretch of a 1-mile trail.

Tanyard Creek is located off of Collier Road, in between Peachtree Road and Northside Drive. There is a concrete trail paved alongside the creek that is part of the Atlanta BeltLine. The trail provides access to open green space and Ardmore Park Playground. Tanyard Creek begins at Tanyard Creek CSO, which is located on Loring Drive. The creek flows downstream into a large open concrete channel that is surrounded by condominiums and greenery. Site 1 is located at the end of the channel, and the sample is collected from the water flowing into the creek. Sites 2 and 3 are located further along the creek, where there is a more natural creek bed. Site 3A is located adjacent to the concrete trail, near Ardmore Park Playground. It is not on the path along sites 2 and 3, and it requires getting on the trail before being able to collect the sample. Site 4 is located further down the creek following the same path as site 3. The first pedestrian bridge, the Beltline trail, and Ardmore park are visible from site 4. Site 5 is located right below the railroad bridge. This is the site of beaver activity, including dams that have been built and destroyed over time. The natural contours of the creek as well as the railroad bridge also cause debris to collect here and slow down the flow of the water. Site 6 is located on the opposite side of the railroad bridge. Around sites 5 and 6 is where the creek is considered Tanyard Creek Park/Ardmore Park. The Beltline trail runs adjacent to the creek from these sites forward. Site 6A is located further down the creek, Sites 7 and 8 are located by the pedestrian bridge where the Beltline trail crosses over Tanyard Creek (map-view included in Appendix). The samples are collected from either side of the bridge, concluding the sample sites. Following the last site, Tanyard Creek merges with several other creeks and flows into the Chattahoochee River.

Following sample collection, membrane filtration (MF) was used to process the samples and analyze the *E. coli* counts. Samples were always collected and processed on the same day, collection was completed. The volume of water that was processed in MF was determined based on water turbidity. Volumes of 0.1ml, 1ml, 5ml, and 10ml were determined in the lab, with higher turbidity requiring less water to be filtered to prevent clogging of the filter. MF technique is as follows: water is passed through the membrane filter, which has a pore size of 0.45um, while the vacuum suction is turned on, filtering all the contents of the water onto the filter paper (sterile, white-grid paper). All contents are then concentrated onto the filter paper, and the filter paper is placed in a petri dish containing BioRad RAPID'*E. coli* 2™ chromogenic medium (Biorad), shown in figure **3.21**. Biorad is a selective chromogenic agar used to count colonies of *E. coli* and other coliforms. The medium detects β-D-Glucuronidase (GLUC), and β-D-Galactosidase (GAL) activities and *E. coli* (GAL+/GLUC+) forms violet to pink colonies, as seen in figure **3.22** (Biorad Laboratories). After filtration, plates are inverted onto the lids, and incubated at 44.5°C for 24 hours. Following incubation, colonies which constituted as *E. coli* colonies were counted as Colony Forming Units (CFU). If colonies were Too Numerous To Count (TNTC), samples were reprocessed at a lower volume and results were collected after the 24-hour incubation period

Samples were adjusted to the standard CFU per 100mL format through the following conversion formula:

$$\Sigma \frac{\boldsymbol{colonies\ per\ site}}{\boldsymbol{volume\ of\ sample\ processed\ per\ site}} \times \mathbf{100} = \underline{\qquad\quad}\ \textbf{CFU per 100 mL}$$

After calculating the CFU per 100mL, the results were normalized by calculating the $\log_{10}$ (CFU per 100mL), and this is how the data is presented and used for statistical analysis.

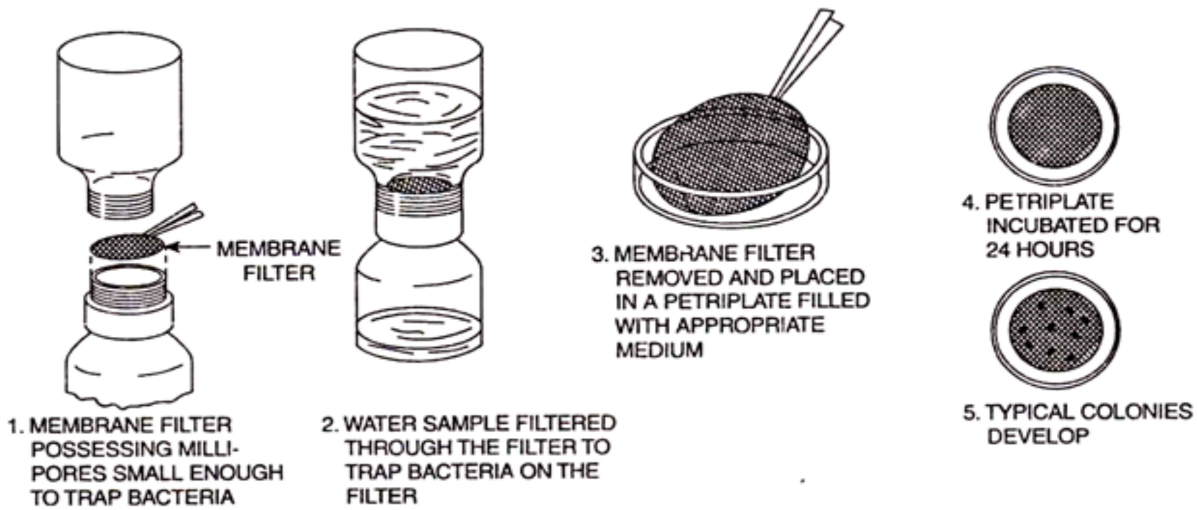


FIG. 19.13. Steps of membrane filter technique (1-5).

Figure 3.21 *Diagram of Membrane Filtration process*

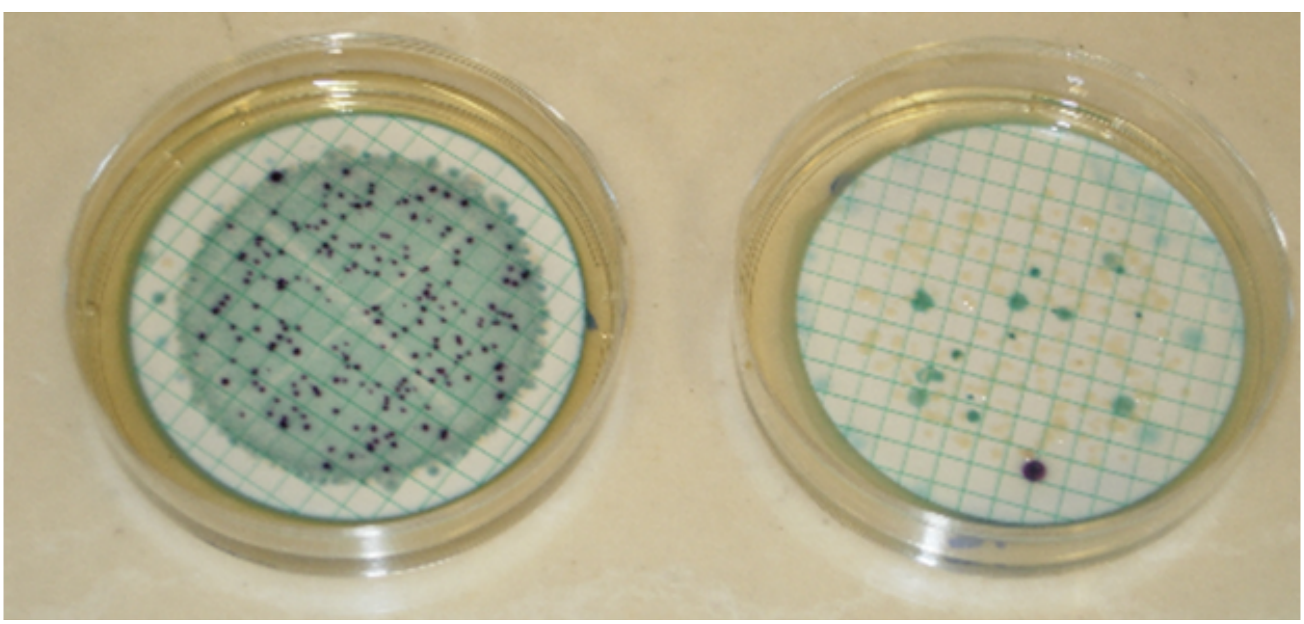


Figure 3.22 *E. coli colonies on Biorad medium agar*

3.2 Secondary Data Collection

Data on the rainfall was collected through the National Weather Service website using the NOWData – NOAA Online Weather Data retrieval search tool for average rainfall amounts. The rainfall was reported in inches. Rainfall data was collected on mean rainfall 24 hours (24h), 48 hours (48h), and a cumulative of 24 and 48 hours (48h cumulative) before sampling. Some of the days had "Trace" amounts of rainfall, which constitutes as not enough rainfall to be measured, but not zero rainfall either. Trace amounts are a small amount of precipitation that will wet a rain gage but is less than the 0.01-inch measuring limit (NOWData).

3.3 Data Analysis

Data was compiled, sorted, organized, and normalized in Microsoft Excel. Microsoft Excel was also used average the *E. coli* count per date as well as the *E. coli* average per site. Graphpad Prism 8 was used to make the graphs, tables, and visuals for the rainfall data. The box and whisker plots provide a visual of the concentration of *E. coli* trends per date as well as per site. The '+' in the box and whisker plot indicate the average of the presented data, and the rest of the plot represents the variability in the data. The dual y-axis graph of *E. coli* concentration and rainfall amount provides a visual representation of both data. In all three of the visuals, the line emerging from Y=2.1 $\log_{10}$ CFU/100mL represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (EPA 2012). The purpose of this line is to indicate how *E. coli* concentrations in Tanyard Creek exceed acceptable recreational water quality standards, demonstrating that the creek is impaired. SAS was used for all tables, graphs and visuals used to model linear and logistic regression. All ANOVA tables from both linear and logistic regression are results from SAS outputs.

3.4 Statistical Analysis

All "Trace" amounts of rainfall were treated as zeros values for rainfall and were in the "no rainfall" category when conducting logistic regression. All rainfall data is broken down into the three rainfall amounts: 24h, 48h and 48h cumulative. All missing values of *E. coli* CFU were excluded from statistical analysis. A majority of the missing values resulted from TNTC values in the primary collection of *E. coli* count. There were also missing values for sites that were not sampled on some days. In addition to that, sample collection was stopped at site 3A after a specific time period (week of 11/18/2019), resulting in the lowest number of N-values for that site.

Graphpad Prism 8 was used to calculate the Spearman Correlation for all the correlation tables and the r^2 values are reported. Spearman correlation used because data was not normal and violated the conditions to conduct a Pearson Correlation. The correlation was calculated between the three rainfall amounts and the *E. coli* concentration of each sampling day for each site. Graphpad Prism 8 was also used to calculate the Mann-Whitney test values when comparing average *E. coli* concentrations per day with and without rain. Rainfall was treated as a binary variable and Mann-Whitney tests were used because the data violated the normalcy assumption of T-tests. Further analysis was conducted comparing *E. coli* concentration per day per site with and without rain for 24h, 48h, and 48h cumulative (in appendix). A statistically significant p-value was alpha=0.05.

SAS was used to conduct all linear and logistic regression modeling and statistical analysis. For both analyses, ANOVA tables were created for average *E. coli* concentration and each rainfall amount. Further analysis was conducted comparing the presence of only rainfall versus all rainfall on the *E. coli* concentration per day for each site (in appendix). To model logistic regression, *E. coli* concentration was a binary outcome: Acceptable vs. not acceptable. The acceptable concentration was $\leq$2.1 $\log_{10}$ CFU, and the not acceptable concentration was >2.1 $\log_{10}$ CFU. A statistically significant p-value was alpha=0.05.

4.1 *E. coli* trends over time and spatially (per site)

Figure 4.11 *E. coli trends spatially (E. coli concentration reported in log_{10} CFU/100mL)*

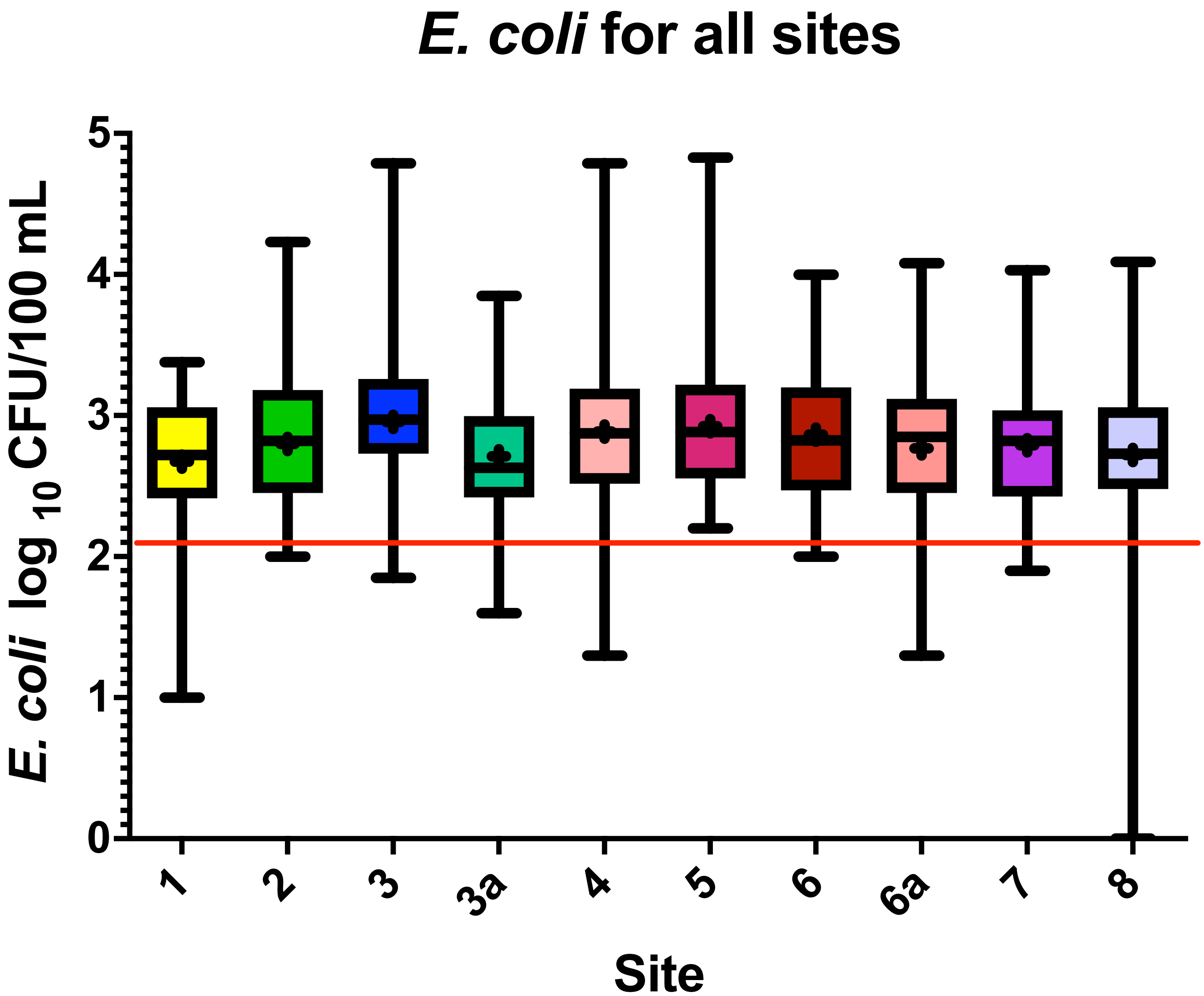


Figure 4.11 The red line represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (2.1 log_{10} CFU/100mL). The + symbols represent the average *E. coli* log_{10} CFU/100mL per site.

Figure 4.12 *E. coli trends over time (E. coli concentration reported in log₁₀ CFU/100mL)*

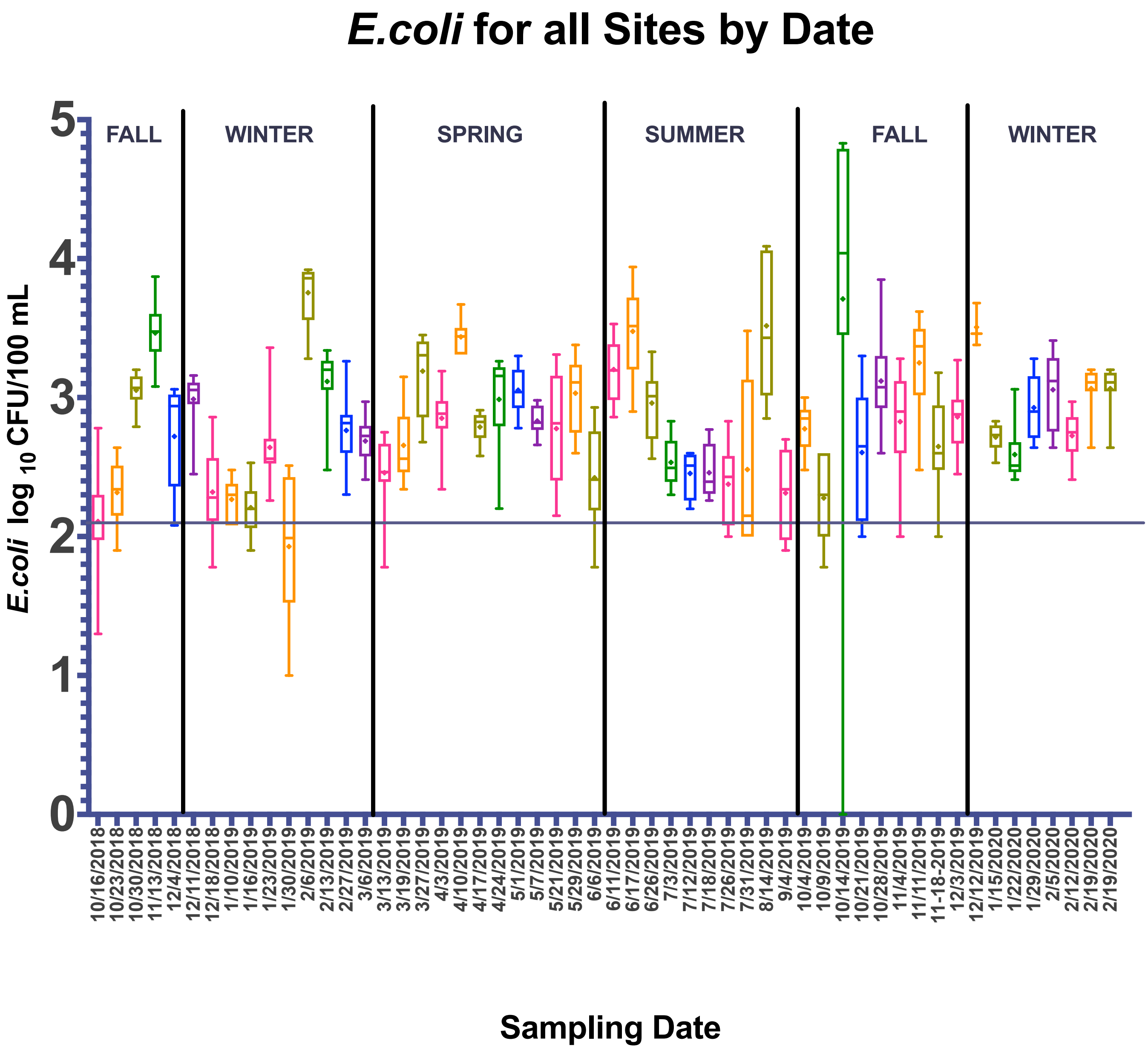


Figure 4.12. The gray line represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (2.1 $\log_{10}$ CFU/100mL). The + symbols represent the average *E. coli* $\log_{10}$ CFU/100mL per sampling date.

<u>Spatial and Temporal Trends:</u> Each sampling site had n-values ranging from 45 to 54, and the average *E. coli* log_{10} CFU/100mL for all sites exceeds the recreational water quality standards **(Figure 4.11)**. There is little variability in the average values of *E. coli* per site (2.5-3 log_{10} CFU/100mL). However, there is variability in the range of *E. coli* concentration per site (4.9 at site 3A and 0 at site 8). The variability of *E. coli* can also be seen in the concentration of *E. coli* per sampling date at Tanyard Creek for all the sites **(Figure 4.12)**. There is variability in the *E. coli* concentration on a weekly basis, and the only consistent aspect is the fact that majority of the mean and median values are above acceptable EPA standards.

4.2 The influence of Rainfall on *E. coli* concentration.

Figure 4.21 *E. coli average for all sampling dates and rainfall 24 hours (24h) prior to sampling*

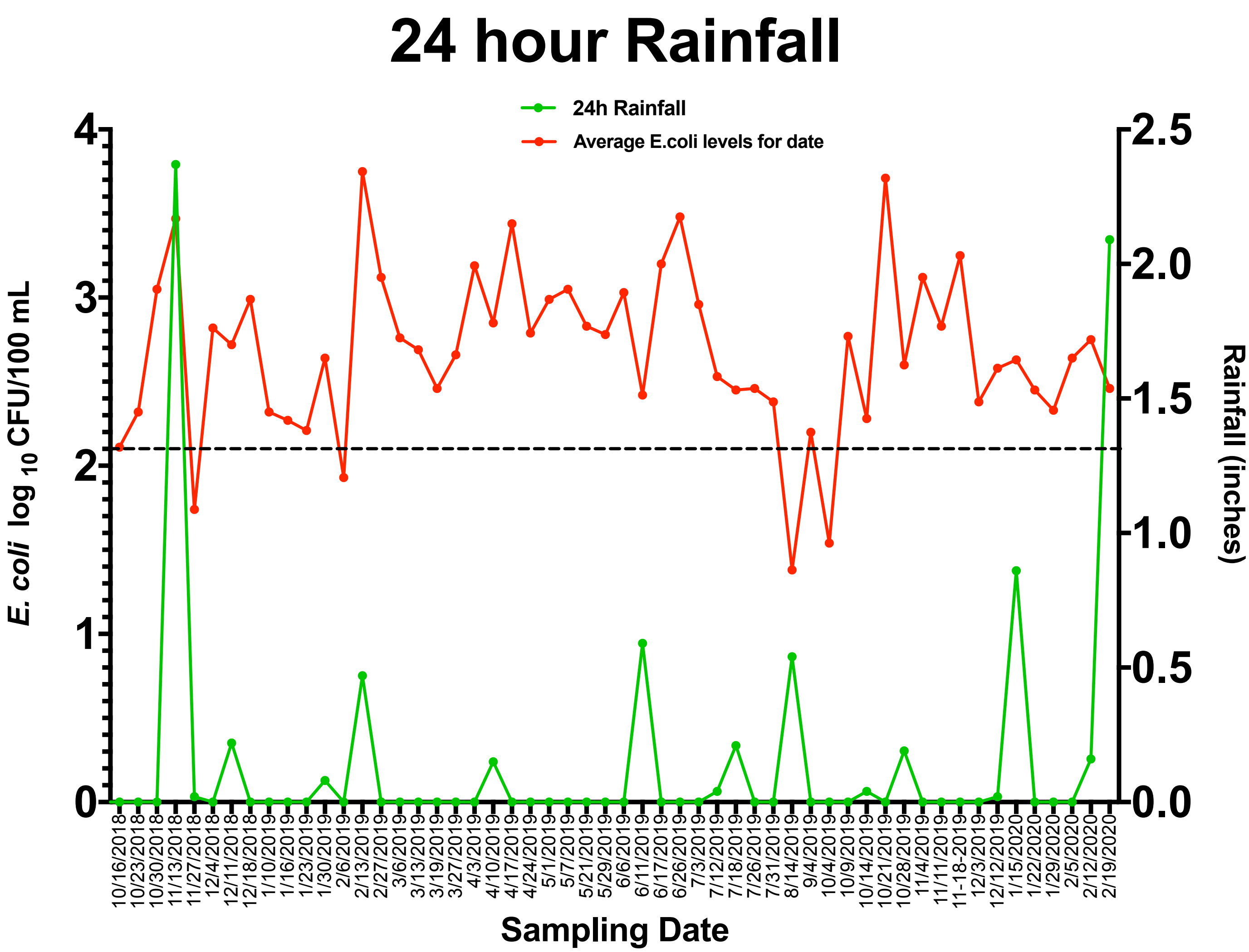


Figure 4.21 The black dotted line represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (2.1 log_{10} CFU/100mL). The red line represents average *E. coli* concentration per date, and the green line represents rainfall in inches 24 hours prior to sampling.

Figure 4.21b *Scatterplot and line of best fit for E. coli average for all sampling dates and rainfall 24h prior to sampling*

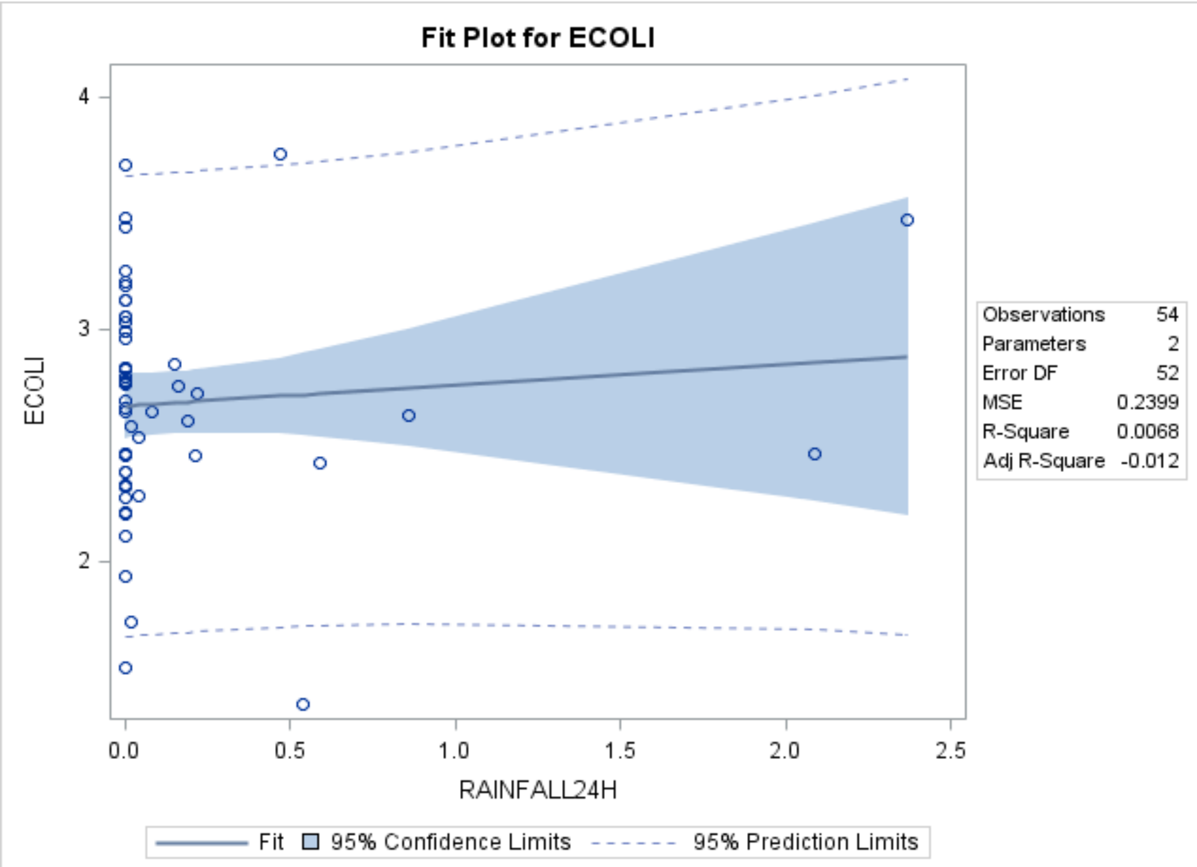


Figure 4.21b is a scatterplot demonstrating the lack of relationship between rainfall 24h prior to sampling and the average concentration of *E. coli* per sampling date.

Table 4.21c *Linear Regression ANOVA table for E. coli average for all sampling dates and rainfall 24h prior to sampling*

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.08483	0.8483	0.35	0.5546
Error	52	12.47414	0.23989		
Corrected Total	53	12.55897			

Roost MSE	0.48978	R-Square	0.0068
Dependent Mean	2.68074	Adj-R-Sq	-0.0123
Coeff Var	18.27043		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.66739	0.07033	37.93	<.0001
Rainfall 24h	1	0.08958	0.15063	0.59	**0.5546**

Table 4.21c shows the linear regression summary when modeling the outcome: average *E. coli* concentration with the predictor: rainfall 24h prior to sampling.

Table 4.21d *Logistic Regression ANOVA table for E. coli average for all sampling dates and rainfall 24h prior to sampling*

Analysis of Maximum Likelihood Estimates					
Parameter	Df	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.5182	0.5473	21.1701	<.0001
Rainfall 24h	1	0.0520	1.2216	0.0018	**0.9660**

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Rainfall 24h	1.053	0.096	11.547

Table 4.21d shows the logistic regression summary when modeling the outcome: average *E. coli* concentration above the acceptable EPA standards (2.1 $\log_{10}$ CFU/100mL) with the predictor: rainfall 24h prior to sampling.

Figure 4.22 *E. coli average for all sampling dates and rainfall 48 hours (48h) prior to sampling*

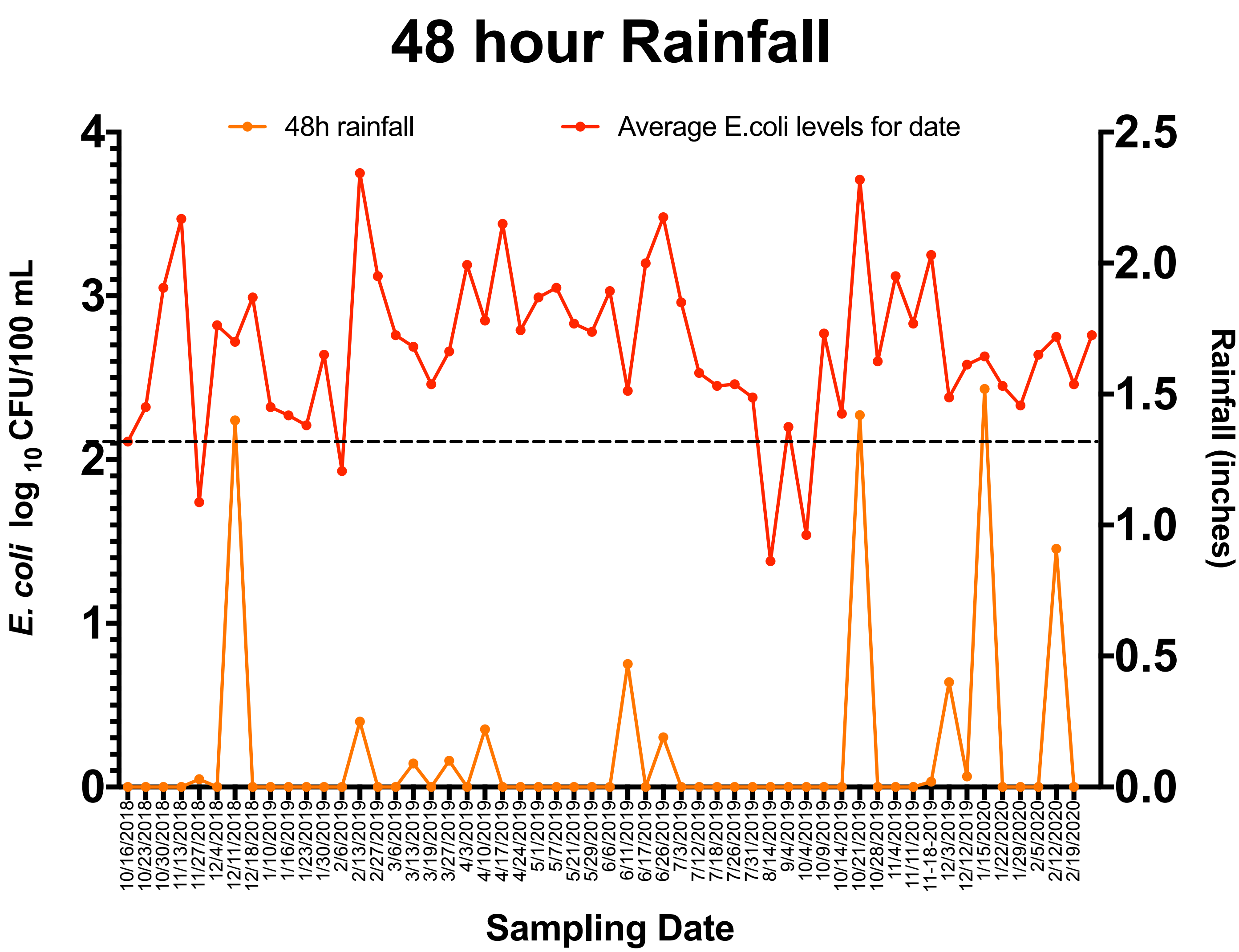


Figure 4.22 The black dotted line represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (2.1 $\log_{10}$ CFU/100mL). The red line represents average *E. coli* concentration per date, and the orange line represents rainfall in inches 48 hours prior to sampling

Figure 4.22b *Scatterplot and line of best fit for E. coli average for all sampling dates and rainfall 48h prior to sampling*

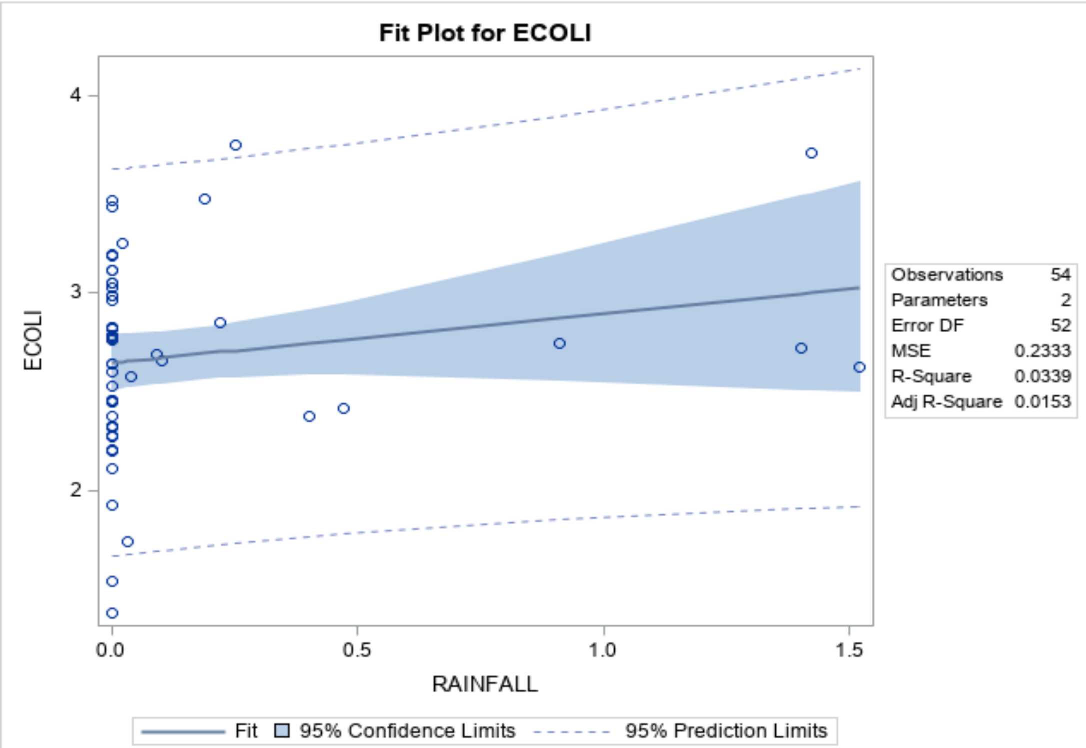


Figure 4.22b is a scatterplot demonstrating the lack of relationship between rainfall 48h prior to sampling and the average concentration of *E. coli* per sampling date.

Table 4.22c *Linear Regression ANOVA table for E. coli average for all sampling dates and rainfall 48h prior to sampling*

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.42565	0.42565	1.82	0.1827
Error	52	12.13332	0.23333		
Corrected Total	53	12.55897			

Roost MSE	0.48305	R-Square	0.0339
Dependent Mean	2.68074	Adj-R-Sq	0.0153
Coeff Var	18.01911		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.64789	0.07009	37.78	<.0001
Rainfall 48h	1	0.25129	0.18605	1.35	**0.1827**

Table 4.22c shows the linear regression summary when modeling the outcome: average *E. coli* concentration with the predictor: rainfall 48h prior to sampling.

Table 4.22d *Logistic Regression ANOVA table for E. coli average for all sampling dates and rainfall 48h prior to sampling*

Analysis of Maximum Likelihood Estimates					
Parameter	Df	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.3054	0.5362	18.4854	<.0001
Rainfall 48h	1	7.7392	14.6878	0.2776	**0.5983**

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Rainfall 48h	>999.999	<0.0001	>999.999

Table 4.22d shows the logistic regression summary when modeling the outcome: average *E. coli* concentration above the acceptable EPA standards (2.1 $\log_{10}$ CFU/100mL) with the predictor: rainfall 48h prior to sampling.

Figure 4.23 *E. coli average for all sampling dates and rainfall 24 hours + 48 hours (48 hours cumulative: 48hcl) prior to sampling*

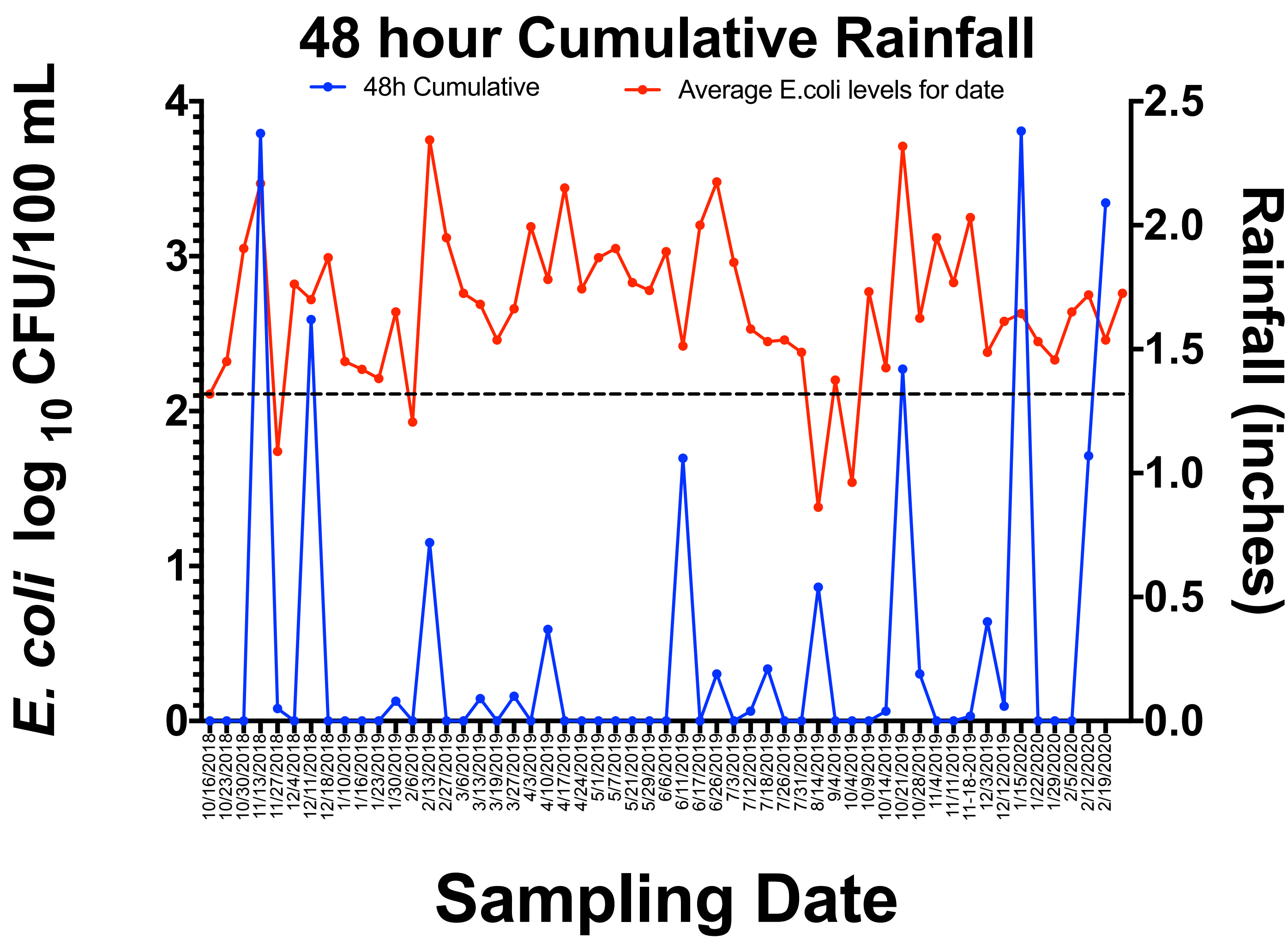


Figure 4.23 The black dotted line represents the acceptable geometric mean of *E. coli* that can be present in recreational waters (2.1 $\log_{10}$ CFU/100mL). The red line represents average *E. coli* concentration per date, and the blue line represents rainfall in inches 48 hours cumulative prior to sampling.

Figure 4.23b *Scatterplot and line of best fit for E. coli average for all sampling dates and rainfall 48hcl prior to sampling*

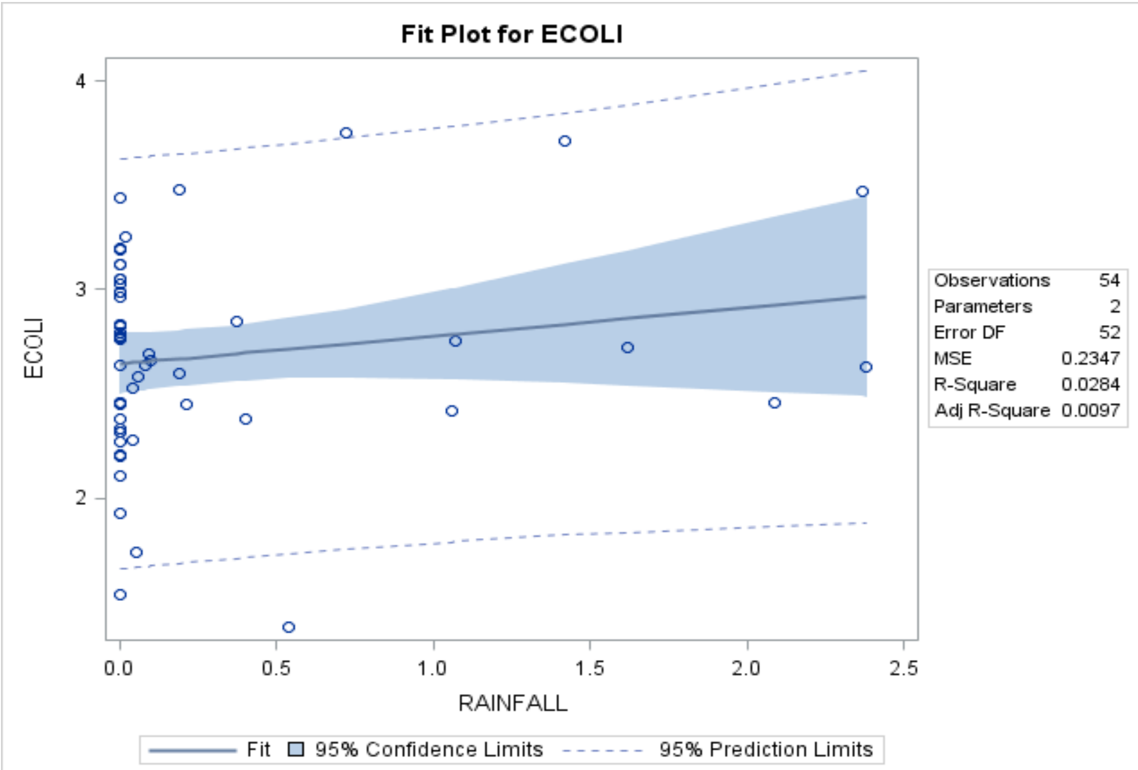


Figure 4.23b is a scatterplot demonstrating the lack of relationship between rainfall 48hcl prior to sampling and the average concentration of *E. coli* per sampling date.

Table 4.23c *Linear Regression ANOVA table for E. coli average for all sampling dates and rainfall 48hcl prior to sampling*

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.35605	0.35605	1.52	0.2236
Error	52	12.20292	0.23467		
Corrected Total	53	12.55897			

Roost MSE	0.48443	R-Square	0.0284
Dependent Mean	2.68074	Adj-R-Sq	0.0097
Coeff Var	18.07072		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.64302	0.07269	36.36	<0.0001
Rainfall 48hcl	1	0.13482	0.10945	1.23	**0.2236**

Table 4.23c shows the linear regression summary when modeling the outcome: average *E. coli* concentration with the predictor: rainfall 48hcl prior to sampling.

Table 4.23d *Logistic Regression ANOVA table for E. coli average for all sampling dates and rainfall 48hcl prior to sampling*

Analysis of Maximum Likelihood Estimates					
Parameter	Df	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.4092	0.5550	18.8407	<.0001
Rainfall 48h Cuml	1	0.5571	1.2610	0.1952	**0.6587**

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Rainfall 48h	1.647	0.147	20.668

Table 4.23d shows the logistic regression summary when modeling the outcome: average *E. coli* concentration above the acceptable EPA standards (2.1 $\log_{10}$ CFU/100mL) with the predictor: rainfall 48hcl prior to sampling.

Rainfall and *E. coli:* Based on **Figure 4.21**, there is no visual relationship between 24h rainfall and elevated *E. coli* concentrations. This is supported by statistical analyses in **Table 4.21c** and **Table 4.21d**, which concludes there is not a statistically significant relationship at alpha=0.05 between rainfall 24h prior to sampling and elevated *E. coli* concentration (p=0.5546 and p=0.9660 respectively). A similar pattern is seen with rainfall 48h prior to sampling, based on **Figure 4.22**, there is no visual relationship between 48h rainfall and elevated *E. coli* concentrations. This is also supported by statistical analyses in **Table 4.22c** and **Table 4.22d**, which concludes there is not a statistically significant relationship at alpha=0.05 between rainfall 48h prior to sampling and elevated *E. coli* concentration (p=0.1827 and p=0.5983 respectively). Finally, at rainfall 48h cumulative, there is no visual relationship between rainfall and elevated *E. coli* concentrations **(Figure 4.23)**. Statistical analyses in **Table 4.23c** and **Table 4.23d** supports the conclusion that there is not a statistically significant relationship at alpha=0.05 between rainfall 48h cumulative prior to sampling and elevated *E. coli* concentration (p=0.2236 and p=0.6587 respectively).

Table 4.24 *Comparing E. coli average for all sampling dates with and without rain*

Average *E.coli* per sampling day	**P-Value**	**Statistically significant at alpha=0.05?**	**Comments**
24h rainfall VS 24h no rainfall	0.2627	No	There were only 16 values present for rainfall, while no rainfall had 38 values.
48h rainfall VS 48h no rainfall	0.3751	No	There were only 14 values present for rainfall, while no rainfall had 43 values
48h cumulative rainfall VS 48h cumulative no rainfall	0.9826	No	There were 22 values present for rainfall, while no rainfall had 32 values

Table 4.24 shows the results of the Mann-Whitney Test when comparing *E. coli* average for all sampling dates with and without the presence of rain.

<u>*E. coli* average for all sampling dates with and without rain:</u> Based on **Table 4.24** there is not a statistically significant relationship at alpha=0.05 between the presence of rain and elevated *E. coli* concentration. However, the results of the Mann-Whitney Test are skewed because there were more n-values present for no rainfall versus rainfall.

Table 4.25 *Spearman Correlation Coefficient of 24h, 48h and 48h Cumulative Rainfall and E. coli per site.*

Site	24h	48h	48h Cumulative
1	0.1458	0.3846	0.4213
2	0.1135	-0.0066	0.1221
3	0.0463*	0.1912	0.1354*
3a	0.2172+	0.5021+	0.2196+
4	0.4163*	-0.3466*	0.0452*
5	0.3753	-0.0056*	0.1668*
6	0.4472*	-0.1923*	0.2125*
6a	0.3074*	-0.2912*	0.1046*
7	0.2091*	0.0330*	0.2047*
8	0.4123*	-0.3191*	0.1957*

*Missing values present, excluded from the analysis
+: Stopped sampling this site, therefore less n values

Table 4.25 shows the Spearman Correlation Coefficient values (r^2) values while calculating the correlation between *E. coli* concentration per sampling day per site and rainfall.

Spearman Correlation: There is no pattern in the r^2 values between each rainfall amount – it is not consistently increasing or decreasing. There is also no trend in the correlation values varying from site to site. This leads to the conclusion that there is not a statistically significant relationship at alpha=0.05 between rainfall and *E. coli* concentration per site.

Chapter V: Discussion

5.1 Discussion of Research Questions

Tanyard Creek is considered to be impaired because it consistently violates its designated use requirements according to the GA EPD 305(b)/303(d) Draft, 2020. The impairment of the creek is also supported through this study, which mainly focused on the *E. coli* concentrations in the creek. The first research question of this study was: How does *E. coli* concentration vary spatially (from site to site) and over time (October 2018-February 2020) in Tanyard Creek? Spatially, the mean and median *E. coli* values for all the sites are above the acceptable EPA quality standards. The average values range between 2.5 and 3 $\log_{10}$ CFU/100mL. Sampling stopped at site 3A during the week of 11/18/2019; therefore, there are fewer N-values for that site that can be a factor in the lower average. Temporally, there is variability in the *E. coli* concentration on a weekly basis, and there is no particular trend by seasons. There is no pattern from the first winter to the second winter of sampling; the same can be seen for the fall seasons. The only consistent factor in Tanyard Creek is the *E. coli* concentrations are above acceptable recreational water quality standards temporally and spatially.

The second research question of this study was: Are rainfall quantities 24 hours, 48 hours, and 48 hours cumulative prior to sampling associated with elevated *E. coli* concentration in Tanyard Creek? For all three of the rainfall quantities, there was no relationship between rainfall and elevated *E. coli* concentration. In essence, there are days where rainfall quantities were high, but *E. coli* concentrations were low, or rainfall quantities are low/none and *E. coli* concentrations are high, or both are high. The relationship between rainfall and *E. coli* concentration was modeled using linear regression, and there was not a statistically significant relationship at alpha=0.05 for all rainfall quantities. The same was done using logistic regression, and the outcome was acceptable/not acceptable *E. coli* concentration. Similar to linear regression, there was not a statistically significant relationship at alpha=0.05. Mann-Whitney Tests also support the hypothesis and conclude there is no statistically significant relationship between elevated *E. coli* concentration and rainfall. Finally, Spearman's correlation leads to the same conclusion: there is no relationship between rainfall and elevated *E. coli* concentrations per site.

The findings of this study correspond with findings from Kleinheinz et al., where they weren't able to draw specific conclusions about the effect of rainfall on *E. coli* concentrations because the data was not uniform. A similar pattern was seen in the study by Ackerman and Weisberg, where there was not a clear-cut conclusion about the effect of rainfall on *E. coli*. However, the conclusions of this study contrast with the Miami Conservatory District report which concluded that there was a clear relationship between rainfall and *E. coli* concentration. This report concluded wet weather elevates *E. coli* concentration in waterbodies. Moreover, the consistent elevated *E. coli* concentration, and presence of large pollutants in Tanyard Creek coincide with the fact that CSOs contribute to urban water pollution (Lund et al., 2014).

5.2 Study Strengths and Limitations

The strength of this study is the collection of primary data for over a year. This helps determine seasonality and trends (or lack thereof). There was consistency in the methods used to collect and process the samples.

The first limitation of the study is the TNTC *E. coli* colonies. Although there weren't a large number of TNTC values, they had to be excluded from statistical and data analyses. The values would be beneficial in increasing the power of statistical analyses and creating visuals. The second limitation of this study was the “trace” amount of rainfall present when retrieving data on rainfall through NOAA. Trace amounts are a small amount of precipitation; therefore, it depends on the investigator of the study to determine whether to count that as rainfall or not. Finally, the third limitation of the study was the unbalanced N-values for rainfall and no rainfall. When conducting the Mann-Whitney tests, no rainfall had more values than rainfall, which can lead to biased results. Moreover, there were less than 20 values present for rainfall data.

5.3 Interventions and Solution

The easy short-term solution is to prevent Tanyard Creek from being clogged by external pollution. This can be achieved through organized community clean-ups of Tanyard Creek. Environmental non-profits, community groups, high schools, and/or colleges can lead weekly or biweekly clean ups and ensure the creek is sanitary.

The harder long-term solution will require a financial commitment from the City of Atlanta to provide replacement equipment that will reduce water quality violations. This includes water mains, catch basins, pump stations, water meters, and filtering equipment. The City will also have to budget for the equipment and preventative maintenance for future improvements to the CSO facilities. Finally, the City of Atlanta will have to accommodate for more space and remodel parts of the CSO system that are faltering. This includes more CSO storage facilities and providing better treatment discharge before it is released into public waters.

5.4 Future Recommendations

The next step for this project would be to continue to collect data on antibiotic-resistant bacteria present in Tanyard Creek. There is currently not enough data to make evidence-based conclusions. *E. coli* colonies forming on Biorad and MacConkey medium agar with varying concentrations of an antibiotic present. Once there are enough N-values to conduct analysis, the results can provide a clearer picture regarding the water quality present in Tanyard Creek.

Another direction for this project will be to determine if there is a relationship between warmer temperatures and elevated *E. coli* concentration. There could be another study similar to the one done by Kleinheinz et al., where Tanyard Creek can be sampled at varying hours after rainfall to determine whether there is a relationship between rainfall and elevated *E. coli* concentration. This study can also conclude how long after rainfall *E. coli* concentration continues to be elevated.

A final recommendation for this project is to conduct a Time Series Analysis with a compilation of all the Tanyard Creek data. There are approximately three years of data collected that can provide an idea of trends and seasonality in the creek. This analysis can also forecast future trends in *E. coli* in Tanyard Creek.

References

- Ackerman D., Weisberg S. B. (2003). Relationship between rainfall and beach bacterial concentrations on Santa Monica Bay beaches. *Journal of Water & Health, 1*(2),85– 89.
- Boehm, A. B., Ashbolt, N. J., Colford, J. M., Dunbar, L. E., Fleming, L. E., Gold, M. A., Hansel, J. A., Hunter, P. R., Ichida, A. M., McGee, C. D., Soller, J. A., & Weisberg, S. B (2009). A sea change ahead for recreational water quality criteria. *Journal of Water and Health*, *7*(1), 9-20. https://doi.org/10.2166/wh.2009.122
- Lund, A., McMillan, J., Kelly, R., Jabbarzadeh, S., Mead, D.G., Burkot, T.R., Kitron, U., Vazquez-Prokopec, G.M. (2014). Long term impacts of combined sewer overflow remediation on water quality and population dynamics of Culex quinquefasciatus, the main urban West Nile virus vector in Atlanta, GA. *Environmental Research*, *129*, 20-26, https://doi.org/10.1016/j.envres.2013.12.008.
- Benjamin-Chung, J., Arnold, B. F., Wade, T. J., Schiff, K., Griffith, J. F., Dufour, A. P., Weisberg, S. B., Colford, J. M., Jr (2017). Coliphages and Gastrointestinal Illness in Recreational Waters: Pooled Analysis of Six Coastal Beach Cohorts. *Epidemiology*, *28*(5), 644–652. https://doi.org/10.1097/EDE.0000000000000681
- BioRad Laboratories, Inc. Lauer W, Armijo C, Patel A. Evaluation of a Chromogenic Medium for Enumeration and Differentiation of Escherichia coli and Other Coliform Bacteria in Selected Foods. Bulletin 5489. Rev A. Food Science Tech Note 5489. Accessed from [http://www.biorad.com/webroot/web/pdf/lsr/literature/Bulletin_5489.pdf].
- Calhoun, L., Avery, M., Jones, L., Gunarto, K., King, R., Roberts, J., Burkot, T. (2007). Combined Sewage Overflows (CSO) Are Major Urban Breeding Sites for Culex quinquefasciatus in Atlanta, Georgia. *The American Journal of Tropical Medicine and Hygiene. 77* (3), 478-84. DOI:10.4269/ajtmh.2007.77.478.
- Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L. (2015). Urban microbial ecology of a freshwater estuary of Lake Michigan. *Elementa Science Anthropocene, 3*, p.000064. DOI: http://doi.org/10.12952/journal.elementa.000064
- Hogan, J. N., Daniels, M. E., Watson, F. G., Conrad, P. A., Oates, S. C., Miller, M. A., Hardin, D., Byrne, B. A., Dominik, C., Melli, A., Jessup, D. A., & Miller, W. A. (2012). Longitudinal Poisson regression to evaluate the epidemiology of Cryptosporidium, Giardia, and fecal indicator bacteria in coastal California wetlands. *Applied and Environmental Microbiology, 78*(10), 3606–3613. https://doi.org/10.1128/AEM.00578-12
- Kleinheinz, G. T., McDermott, C. M., Hughes, S., & Brown, A. (2009). Effects of Rainfall on E. coli Concentrations at Door County, Wisconsin Beaches. *International Journal of Microbiology,* 876050. https://doi.org/10.1155/2009/876050
- Koirala, S.R., Gentry, R.W., Perfect, E., Schwartz, J.S., Sayler, G.S., (2008). Temporal variation and persistence of bacteria in streams. *Journal of Environmental Quality 37*, 1559–1566.
- Koirala, S. R., Gentry, R. W., Perfect, E., Schwartz, J. S., Sayler, G. S. (2008). Temporal Variation and Persistence of Bacteria in Streams. *Journal of Environmental Quality 37*, 1559-1566. doi:10.2134/jeq2007.0310
- Miami Conservancy District Report. Report No. 2018-23. Relationship of Escherichia Coli Levels to Rainfall, Runoff, and Water Quality Variables at Two Urban Sites in the Great Miami River Watershed MCD Report No. 2018-23

- Rossi, A., Wolde, B. T., Lee, L. H., & Wu, M. (2020). Prediction of recreational water safety using Escherichia coli as an indicator: case study of the Passaic and Pompton rivers, New Jersey. *Science of the Total Environment*, 714, [136814]. https://doi.org/10.1016/j.scitotenv.2020.136814
- Schilling, K., Zhang, Y.K., Hill, D.R., Jones, C.S., Wolter, C.F. (2009). Temporal variations of Escherichia coli concentrations in a large Midwestern river. *Journal of Hydrology, 365*(1-2), 79-85. https://doi.org/10.1016/j.jhydrol.2008.11.029.
- Tibbetts J. (2005). Combined sewer systems: down, dirty, and out of date. *Environmental Health Perspective;113*(7): A464–A467. doi:10.1289/ehp.113-a464
- U.S. Environmental Protection Agency (U.S. EPA) (2012). Recreational Water Quality Criteria. Office of Water 820-F-12-058.
- Georgia Environmental Protection Agency (GA EPD) (2020). Draft 2020 Integrated 305(b)/303(d) List – Streams.

Images and Figures

- **Figure 1.21** Taken from: Performance Audit: Combine Sewage Overflow Consent Decree Impact (January 2014). City Auditors Office, City of Atlanta. **Source:** Cambridge Department of Public Works, Cambridge Massachusetts.
- **Figure 1.31** Taken from: Performance Audit: Combine Sewage Overflow Consent Decree Impact (January 2014). City Auditors Office, City of Atlanta. Source**:** Prepared by city auditor's staff using information from the Department of Watershed Management
- **Membrane Filtration Diagram:** http://www.biologydiscussion.com/microorganisms/measurement-of-cell-numbers-and-cell-mass-of-microorganisms/55160

Appendix

Mann- Whitney Test of comparing *E. coli* levels per day per site with and without rain 24h before sampling.

Site	P-Value	Statistically significant at alpha=0.05?	Comments
Site 1	0.0357	Yes	Missing values were excluded from analysis. There were 16 values for rain present, and 38 values for no rain
Site 2	0.1268	No	Missing values were excluded from analysis. There were 16 values for rain present, and 38 values for no rain
Site 3	0.1947	No	Missing values were excluded from analysis. There were 14 values for rain present, and 37 values for no rain
Site 3a	0.2026	No	Missing values were excluded from analysis. Sampling stopped at site 3a after a certain time period, those zero values were also excluded from analysis. There were 12 values for rain present, and 34 for no rain
Site 4	0.1769	No	Missing values were excluded from analysis. There were 14 values for rain present, and 38 values for no rain
Site 5	0.1342	No	Missing values were excluded from analysis. There were 15 values for rain present, and 39 values for no rain
Site 6	0.1558	No	Missing values were excluded from analysis. There were 15 values for rain present, and 39 values for no rain
Site 6A	0.3403	No	Missing values were excluded from analysis. There were 15 values for rain present, and 39 values for no rain
Site 7	0.1861	No	Missing values were excluded from analysis. There were 15 values for rain present, and 39 values for no rain
Site 8	0.6774	No	Missing values were excluded from analysis. There were 15 values for rain present, and 39 values for no rain

Note: Limitations include not having a balanced number of n values for rainfall vs. no rainfall. Limitations also include not having greater than 20 values for rainfall data. Two-tailed T-test was used to report P-values. Rainfall is treated as a binary variable: No rainfall vs rainfall

Mann- Whitney Test of comparing *E. coli* levels per day per site with and without rain 48h before sampling.

Site	P-Value	Statistically significant at alpha=0.05?	Comments
Site 1	0.2470	No	Missing values were excluded from analysis. There were 14 values for rain present, and 39 values for no rain
Site 2	0.7382	No	Missing values were excluded from analysis. There were 14 values for rain present, and 39 values for no rain
Site 3	0.2734	No	Missing values were excluded from analysis. There were 14 values for rain present, and 37 values for no rain
Site 3a	0.8398	No	No missing values. Sampling stopped at site 3a after a certain time period, those zero values were excluded from analysis. There were 9 values for rain present, and 36 for no rain
Site 4	0.1806	No	Missing values were excluded from analysis. There were 13 values for rain present, and 38 values for no rain
Site 5	0.1695	No	Missing values were excluded from analysis. There were 13 values for rain present, and 40 values for no rain
Site 6	0.1207	No	Missing values were excluded from analysis. There were 13 values for rain present, and 40 values for no rain
Site 6A	0.0084	**Yes**	Missing values were excluded from analysis. There were 13 values for rain present, and 40 values for no rain
Site 7	0.0725	No	Missing values were excluded from analysis. There were 13 values for rain present, and 40 values for no rain
Site 8	0.0341	**Yes**	Missing values were excluded from analysis. There were 13 values for rain present, and 40 values for no rain

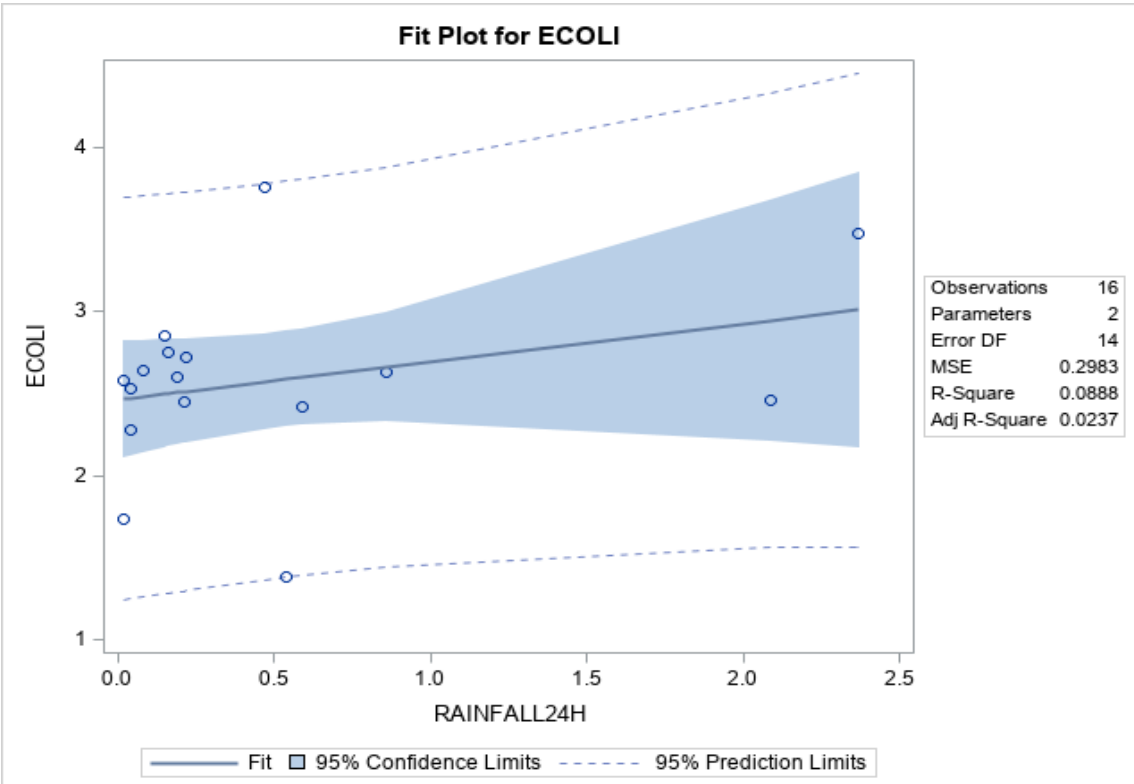
Note: Limitations include not having a balanced number of n values for rainfall vs no rainfall. Limitations also include not having greater than 20 values for rainfall data. Two tailed T-test was used to report P-values. Rainfall is treated as a binary variable: No rainfall vs rainfall

Mann- Whitney Test of comparing *E. coli* levels per day per site with and without rain 48h before sampling

Site	P-Value	Statistically significant at alpha=0.05?	Comments
Site 1	0.2385	No	No missing values. N=22 for rainfall, N=32 for no rainfall
Site 2	0.7301	No	Missing values excluded from analysis. N=22 for rainfall, N=32 for no rainfall
Site 3	0.2112	No	Missing values excluded from analysis. N=20 for rainfall, N=32 for no rainfall
Site 3a	0.2457	No	No missing values. Sampling stopped at site 3a after a certain time period, those zero values were excluded from analysis. N=16 for rainfall, N=29 for rainfall
Site 4	0.0575	No	Missing values excluded from analysis. N=20 for rainfall, N=31 for no rainfall
Site 5	**0.0352**	**Yes**	Missing values excluded from analysis. N=21 for rainfall, N=32 for no rainfall
Site 6	0.0799	No	Missing values excluded from analysis. N=21 for rainfall, N=32 for no rainfall
Site 6A	**0.0278**	**Yes**	Missing values excluded from analysis. N=21 for rainfall, N=32 for no rainfall
Site 7	0.1479	No	Missing values excluded from analysis. N=21 for rainfall, N=32 for no rainfall
Site 8	0.3402	No	Missing values excluded from analysis. N=21 for rainfall, N=32 for no rainfall

Note: Limitations include not having a balanced number of n values for rainfall vs no rainfall. Limitations also include not having greater than 20 values for rainfall data. Two tailed T-test was used to report P-values. Rainfall is treated as a binary variable: No rainfall vs rainfall

Linear Regression Summary: Line of Best Fit and Anova Table for Average *E. coli* 24h before sampling (ONLY rainfall)



ANOVA TABLE

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.40712	0.40712	1.36	0.2622
Error	14	4.17672	0.29834		
Corrected Total	15	4.58384			

Roost MSE	0.54620	R-Square	0.0888
Dependent Mean	2.57813	Adj-R-Sq	0.0237
Coeff Var	21.18604		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.46260	0.16860	14.61	<.0001
Rainfall 24h	1	0.22961	0.19656	1.17	**0.2622**

Linear Regression Summary: Rainfall per site for 24h (Rainfall and No rainfall)

- Y= *E. coli*
- X_1 = 24h Rainfall (including no rainfall)

Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	0.1022	No	53	2.77	Y= 2.57+0.32978(X1)
2	0.1465	No	53	2.17	Y= 2.73 + 0.225(x1)
3	0.2889	No	51	1.15	Y= 2.91+0.176(x1)
3a	0.0583	No	45	3.78	Y=2.63 + 0.429(x1)
4	0.0559	No	51	3.84	Y=2.79+ 0.352 (x1)
5	0.1421	No	53	2.22	Y=2.88+0.250 (x1)
6	0.0762	No	53	3.27	Y= 2.74+0.360(x1)
6a	0.2689	No	53	1.25	Y= 2.56+0.285(x1)
7	0.3687	No	53	0.82	Y=2.57+0.219(x1)
8	0.4635	No	53	0.55	Y=2.50+0.194(x1)

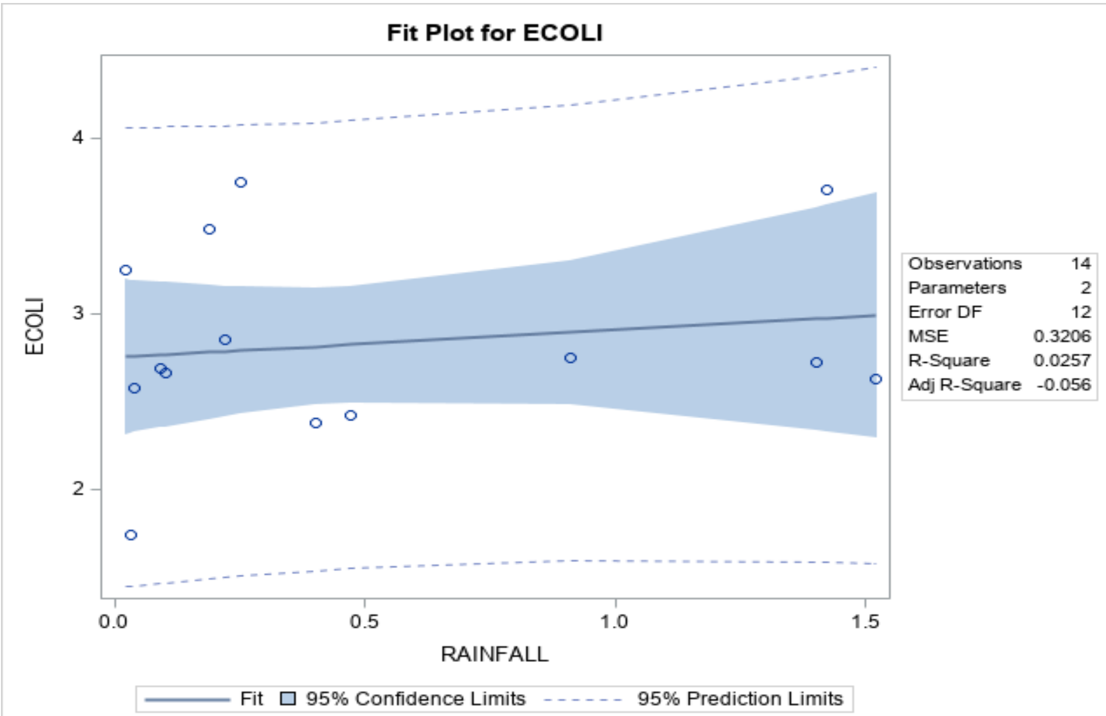
Linear Regression Summary: Rainfall per site for 24h (Rainfall only)

- Y= *E. coli*
- X_1 = 24h Rainfall (only rainfall)

Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	0.3221	No	16	1.05	Y= 2.69+0.216(x1)
2	0.4608	No	16	0.58	Y=2.803+ 0.174(x1)
3	0.6879	No	14	0.17	Y= 2.99+0.120 (x1)
3a	0.3683	No	12	0.89	Y = 2.68+0.389(x1)
4	0.3024	No	14	1.16	Y= 2.83+ 0.329 (x1)
5	0.6366	No	15	0.23	Y=3.06+0.125(x1)
6	0.2484	No	15	1.46	Y=2.86+ 0.279 (x1)
6a	0.3991	No	15	0.76	Y=2.53+0.302(x1)
7	0.5430	No	15	0.39	Y=2.58+0.212(x1)

8	0.3764	No	15	0.84	Y=2.25+0.366(x1)

Linear Regression Summary: Line of Best Fit and Anova Table for Average *E. coli* 48h before sampling (ONLY rainfall)



ANOVA TABLE

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.10143	0.10143	0.32	0.5841
Error	12	3.84687	0.32057		
Corrected Total	13	3.94829			

Roost MSE	0.56619	R-Square	0.0257
Dependent Mean	2.82929	Adj-R-Sq	-0.0555
Coeff Var	20.01180		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.75000	0.20680	13.30	<.0001
Rainfall 48h	1	0.15722	0.27950	0.56	**0.5841**

Linear Regression Summary: Rainfall per site for 48h (Rainfall and no rainfall)

- Y= *E.coli*
- X_1 = 48h Rainfall (including no rainfall)

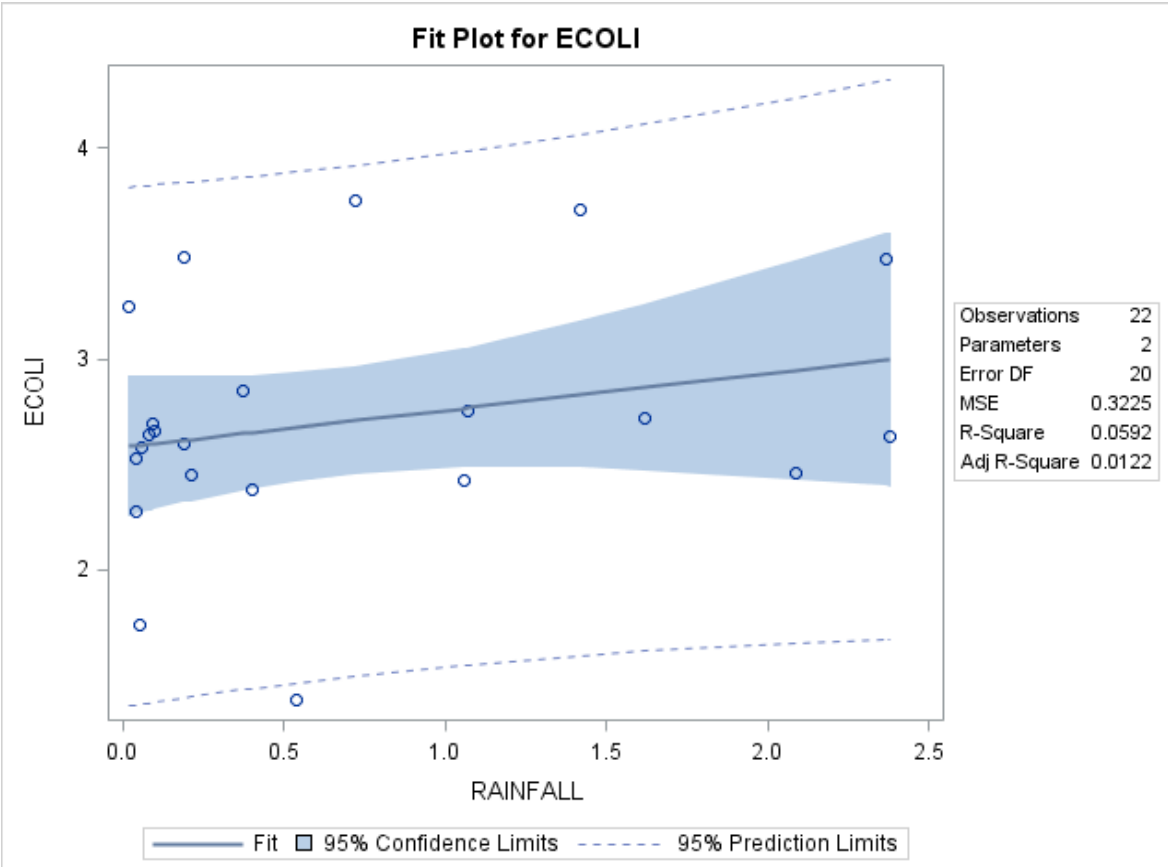
Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	0.1803	No	53	1.85	Y=2.61+0.260(x1)
2	0.4998	No	53	0.46	Y=2.78-0.132(x1)
3	0.7549	No	51	0.10	Y=2.92+0.065(x1)
3a	0.7633	No	45	0.09	Y=2.67+0.086(x1)
4	0.6519	No	51	0.21	Y=2.86-0.105(x1)
5	0.6945	No	53	0.16	Y=2.91+0.084(x1)
6	0.8044	No	53	0.06	Y=2.79+0.064(x1)
6a	0.3153	No	53	1.03	Y=2.56+0.324(x1)
7	0.4173	No	53	0.67	Y=2.57+0.247(X1)
8	0.5599	No	53	0.34	Y=2.50+0.194(x1)

Linear Regression Summary: Rainfall per site for 48h (Rainfall only)

- Y= E.coli
- X_1 = 48h Rainfall (only rainfall)

Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	0.3809	No	13	0.83	Y=2.73+0.151(x1)
2	0.4575	No	13	0.59	Y=2.81-0.154(X1)
3	0.5422	No	13	0.39	Y=3.10-0.098(x1)
3a	0.6495	No	8	0.23	Y=2.58+0.176(X1)
4	**0.0415**	**Yes**	**12**	**5.32**	**Y=3.22-0.434(x1)**
5	0.6912	No	12	0.17	Y=3.09-0.073(x1)
6	0.1681	No	12	2.18	Y=3.15-0.262(x1)
6a	0.1713	No	12	2.14	Y=3.10-0.171(X1)
7	0.5407	No	12	0.40	Y=2.99-0.132(X1)
8	0.0842	No	12	3.60	Y=3.13-0.372(x1)

Linear Regression Summary: Line of Best Fit and Anova Table for Average *E. coli* 48h cumulative before sampling (ONLY rainfall)



ANOVA TABLE

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F-value	Pr > F
Model	1	0.40613	0.40613	1.26	0.2751
Error	20	6.45045	0.32252		
Corrected Total	21	6.85658			

Roost MSE	0.56791	R-Square	0.0592
Dependent Mean	2.70091	Adj-R-Sq	0.0122
Coeff Var	21.02665		

Parameter Estimates					
Variable	Df	Parameter Estimate	Standard Error	T-Value	Pr > |t|
Intercept	1	2.58151	0.16119	16.02	<.0001
Rainfall 48h cuml	1	0.17384	0.15492	1.12	**0.2751**

Linear Regression Summary: Rainfall per site for 48h Cumulative (Rainfall and No rainfall)

- Y= E.coli
- X_1 = 48h cumulative Rainfall

Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	**0.0338**	**Yes**	**53**	**4.76**	**Y=2.57+0.238(x1)**
2	0.5073	No	53	0.45	Y=2.74+0.076(x1)
3	0.3373	No	51	0.94	Y=2.90+0.117(x1)
3a	0.1069	No	45	2.71	Y=2.62+0.278(x1)
4	0.2648	No	51	1.27	Y=2.80+0.152(x1)
5	0.1906	No	53	1.76	Y=2.87+0.164(x1)
6	0.1490	No	53	2.15	Y=2.74+0.216(x1)
6a	0.1591	No	53	2.04	Y=2.53+0.266(X1)
7	0.2543	No	53	1.33	Y=2.55+0.203(X1)
8	0.3776	No	53	0.79	Y=2.48+0.172(X1)

Linear Regression Summary: Rainfall per site for 48h cumulative (Rainfall only)

- Y= E.coli
- X_1 = 48h cumulative Rainfall

Site	P-value	Statistically Significant at alpha=0.05?	N	F-value	Fitted Model
1	0.0934	No	22	3.10	Y=2.58+0.236(X1)
2	0.8295	No	22	0.05	Y=2.82+0.030(X1)
3	0.9225	No	20	0.01	Y=3.06+0.016(X1)
3a	0.4292	No	16	0.66	Y=2.72+0.207(X1)
4	0.8953	No	20	0.02	Y=3.00+0.027(X1)
5	0.9964	No	21	0.00	Y=3.14-0.00073(x1)
6	0.6199	No	21	0.25	Y=2.96+0.078(x1)
6a	0.5695	No	21	0.34	Y=2.75+0.128(X1)
7	0.4876	No	21	0.50	Y=2.63+0.155(X1)
8	0.4819	No	21	0.51	Y=2.45+0.191(X1)

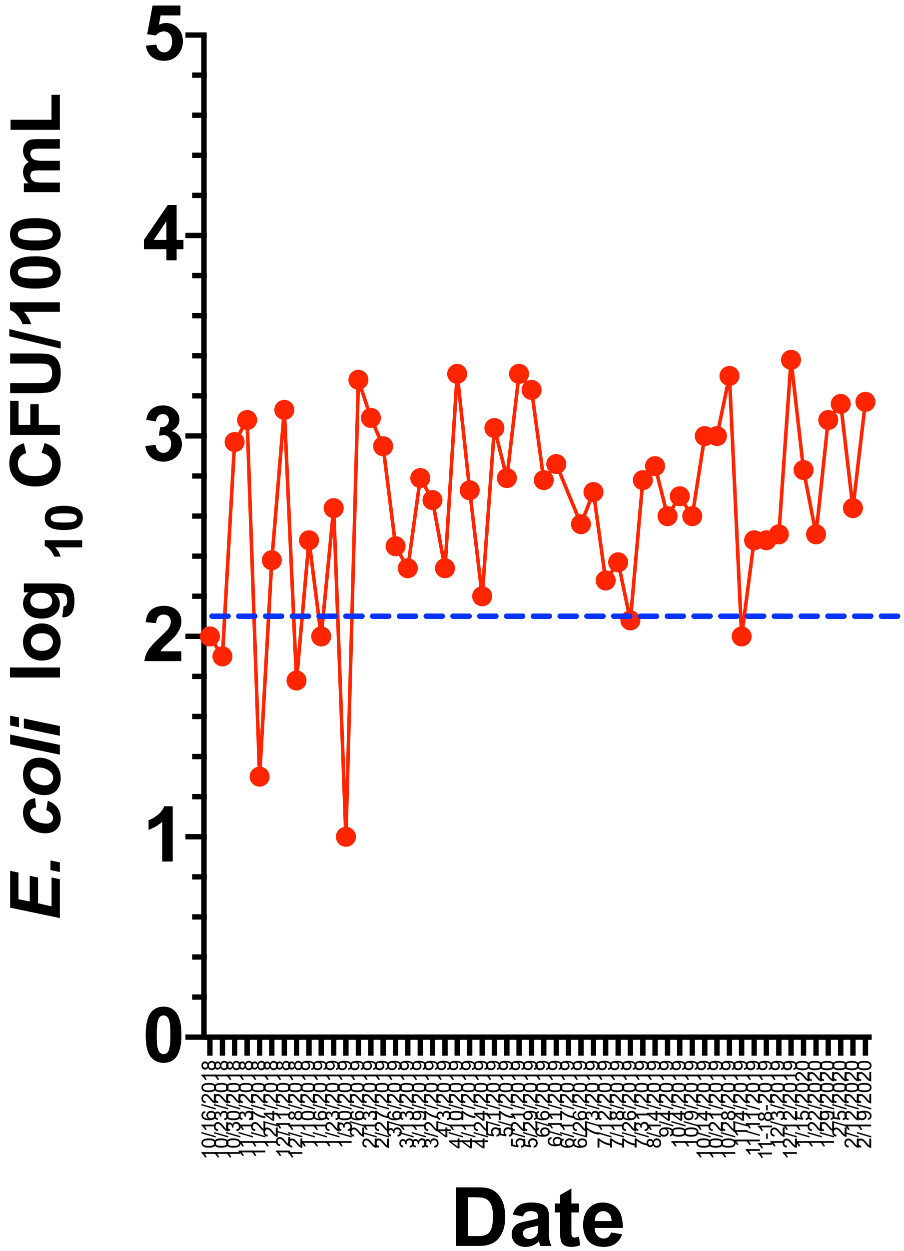
Map of Sampling Sites (Approximation based on map, not exact location)


Site 6
Site 5
Site 4
Site 3A
Northside Beltline Trail
Northside BeltLine Trail
Taynard Creek Trail
Ardmore Park Playground
Google


Site 6A
Northside Beltline Trail
Google
Map data ©2020 , Map data ©2020 United States Terms Send feedback 20 ft

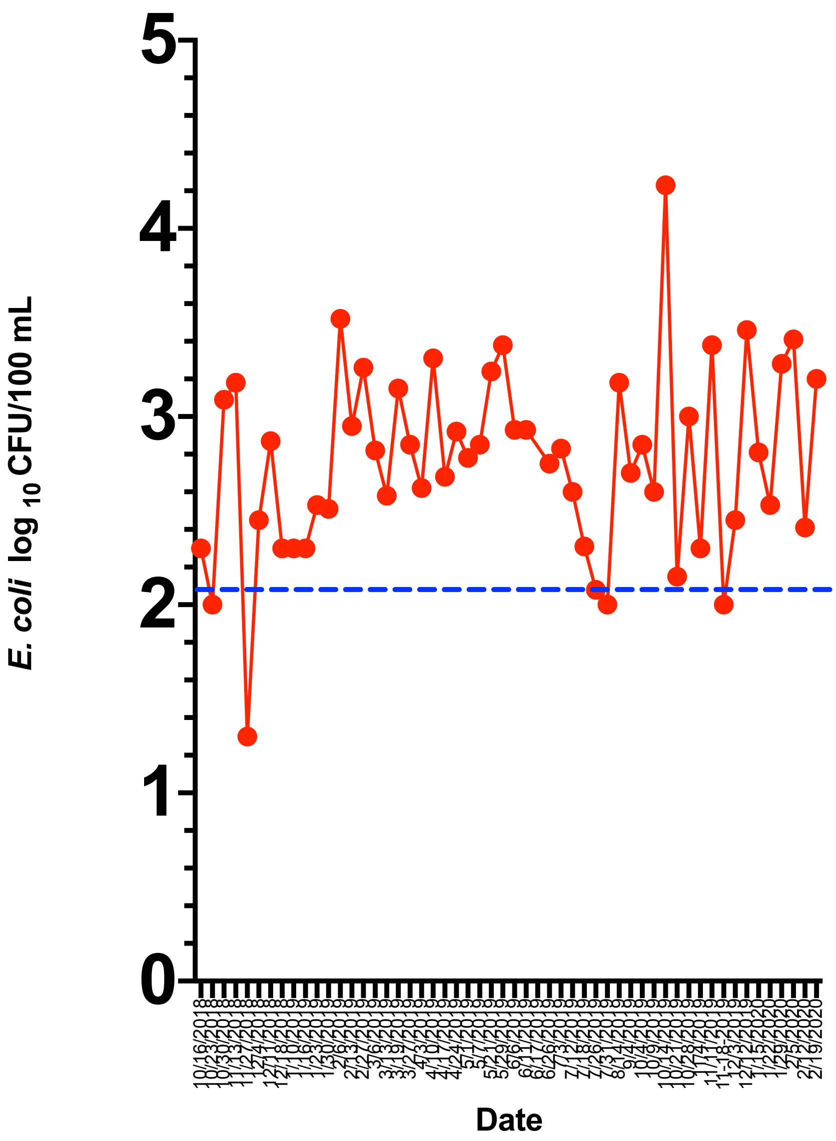

Site 8
Northside Beltline Trail
Northside Beltline Trail
Northside Beltline Trail
Northside Beltline Trail
Site 7
Google

E.coli per date - Site 1



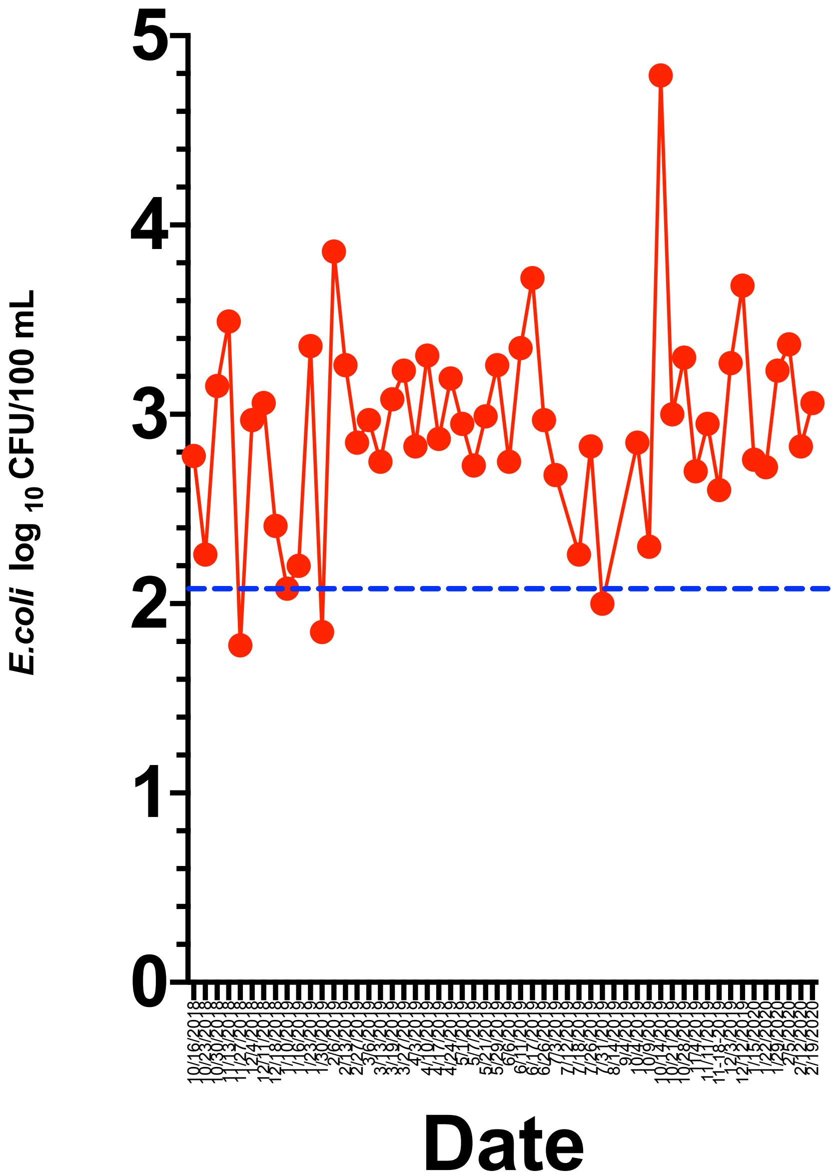
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 2



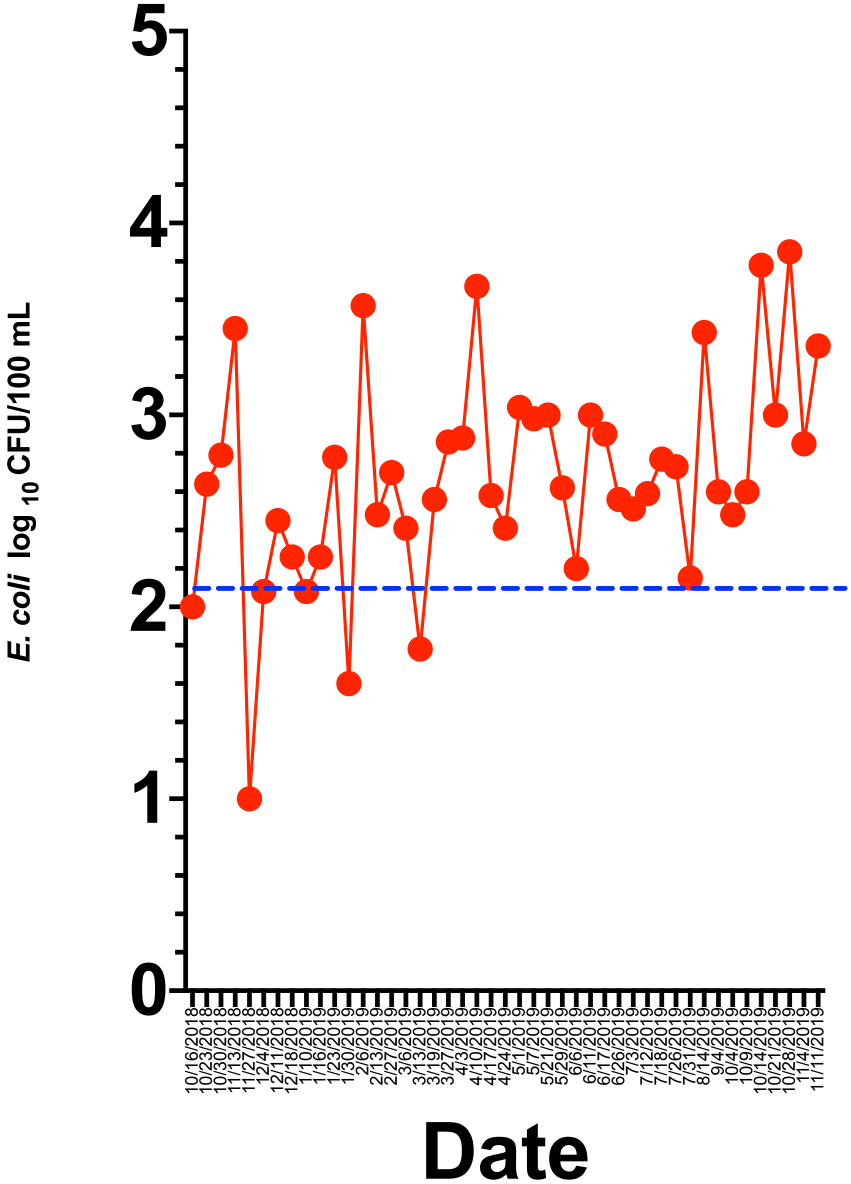
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 3



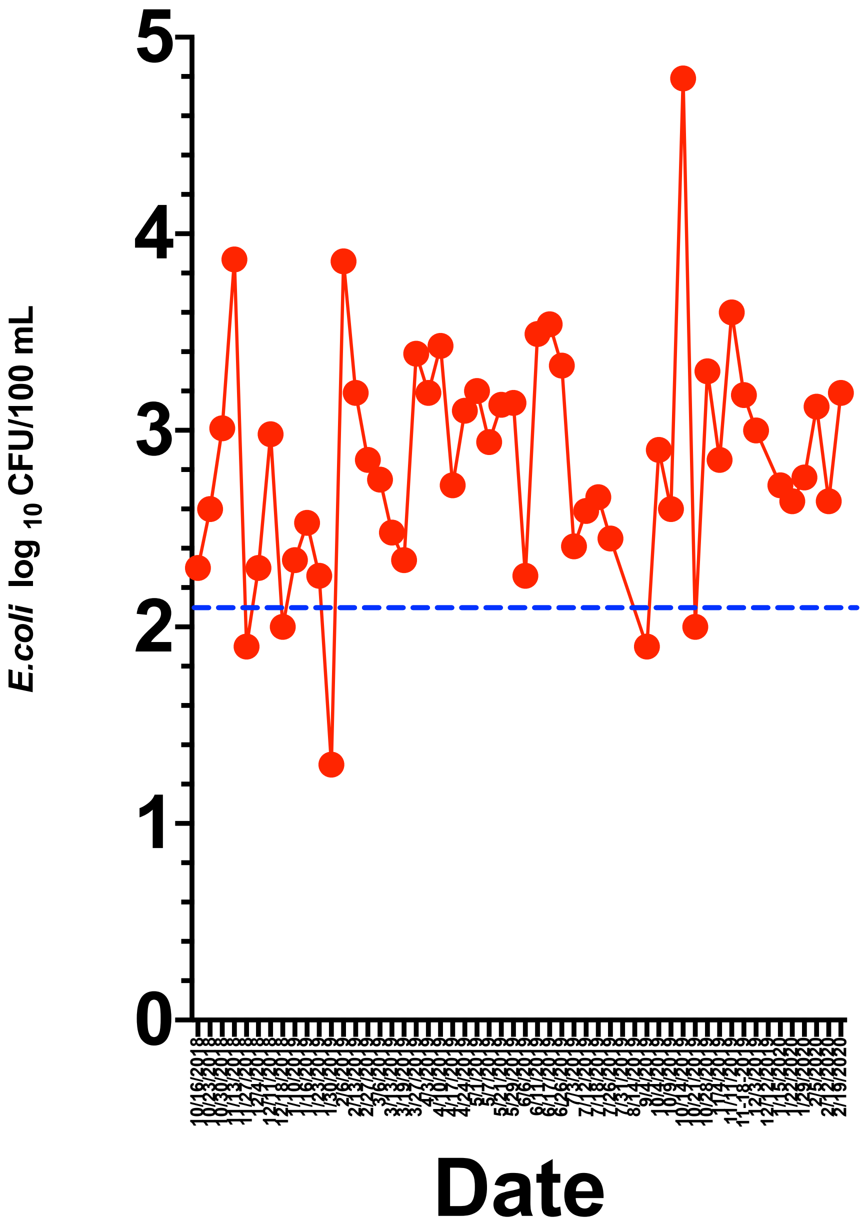
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 log_{10} CFU/100mL, which is represented by the blue line.

E.coli per date- Site 3a



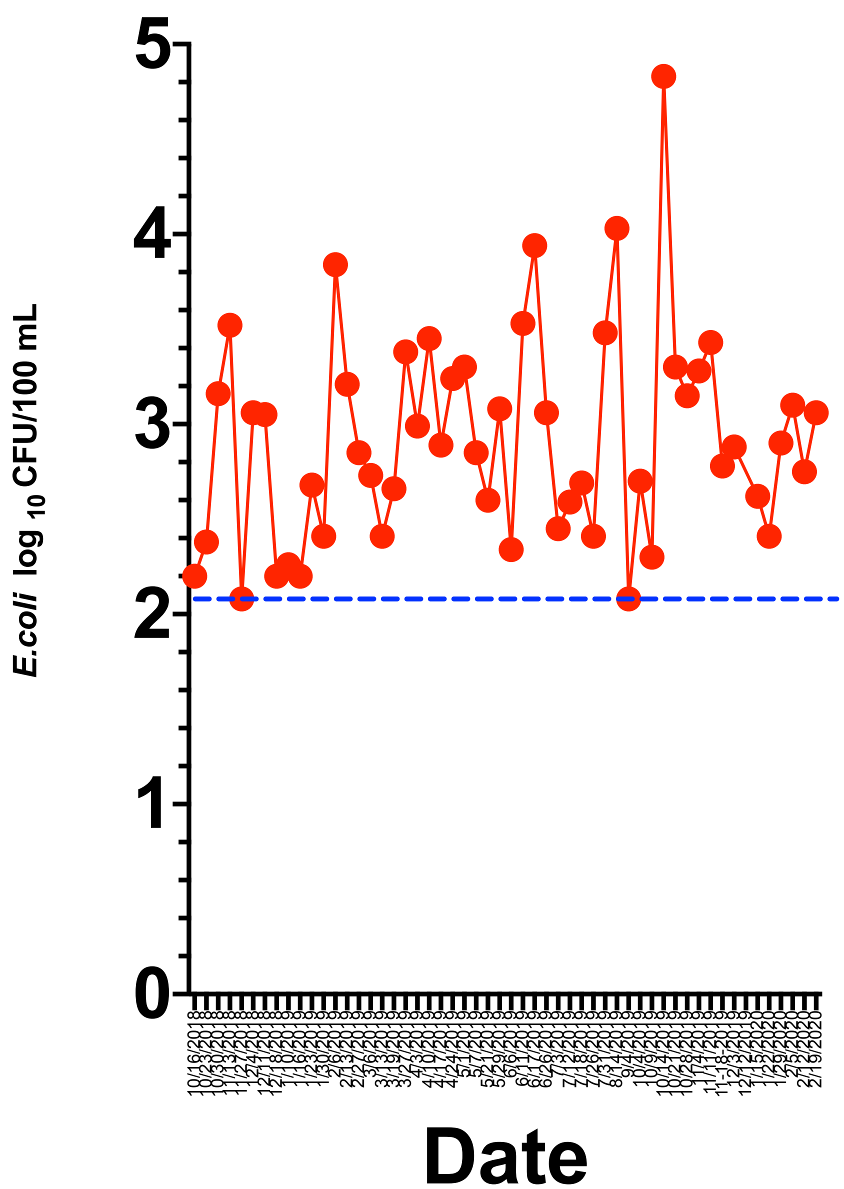
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 4



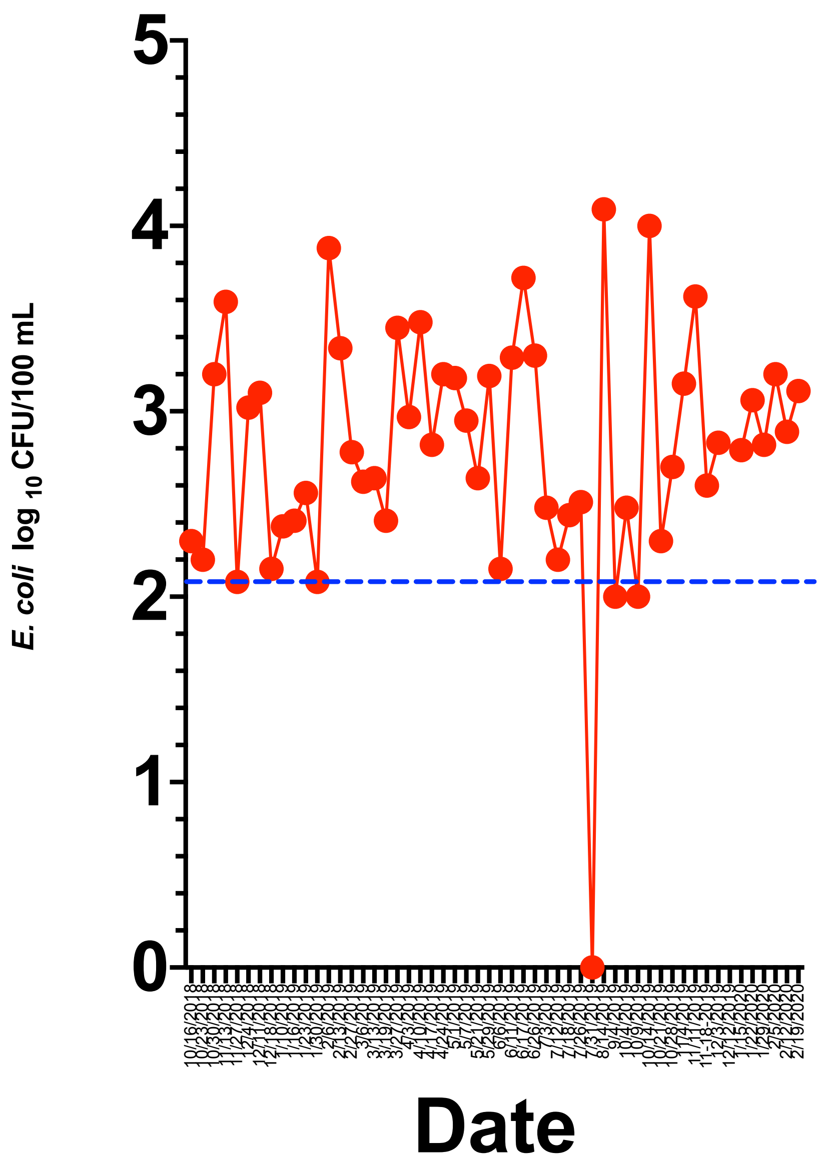
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 5



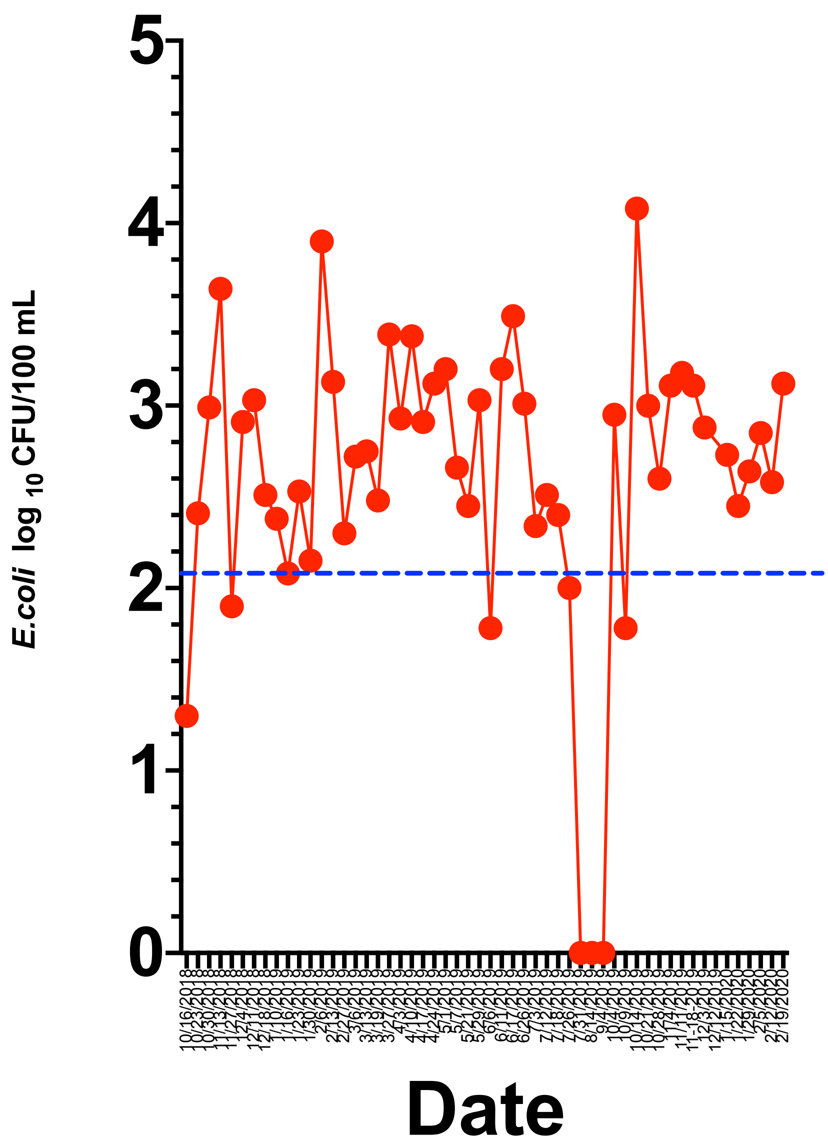
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 6



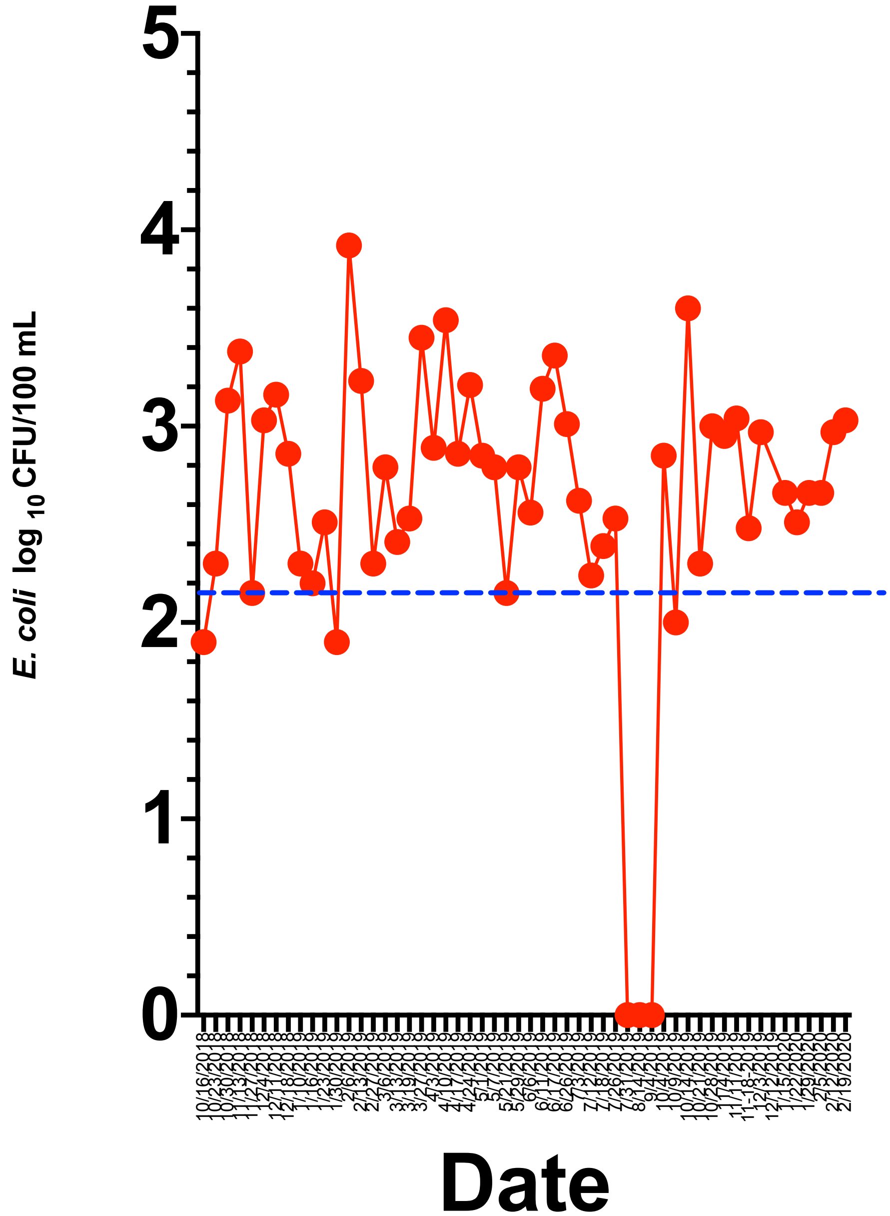
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 log_{10} CFU/100mL, which is represented by the blue line.

E.coli per date-Site 6a



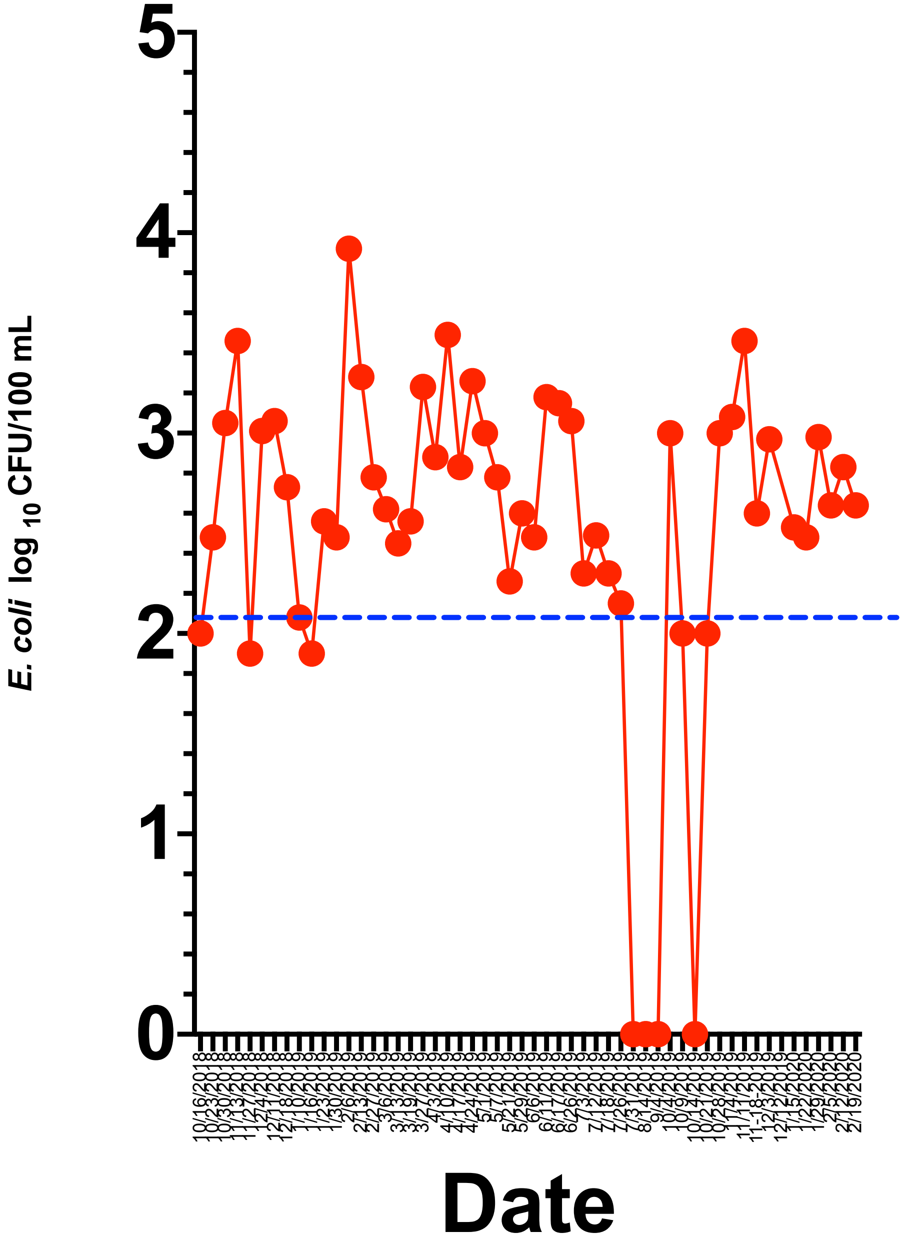
According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 log_{10} CFU/100mL, which is represented by the blue line.

E.coli per date- Site 7



According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.

E.coli per date - Site 8



According to EPA standards, the acceptable geometric mean of *E. coli* that can be present in recreational waters is 126 CFU/100mL or 2.1 $\log_{10}$ CFU/100mL, which is represented by the blue line.